Production of terpenes and terpenoids

Simonsen, Henrik Toft; Peramuna, Anantha Vithakshana

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 March 2019 (28.03.2019)

W 1P O | PCT

(10) International Publication Number
WO 2019/057258 Al

(51) International Patent Classification:
C12N 9/88 (2006.01) *C12P 5/00* (2006.01)
C12N 15/82 (2006.01) *A01H 11/00* (2006.01)
C12P 7/04 (2006.01)

(21) International Application Number:
PCT/DK20 18/05023 1

(22) International Filing Date:
2 1 September 2018 (21.09.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
17192650.4 22 September 2017 (22.09.2017) EP

(71) Applicant: DANMARKS TEKNISKE UNIVERSITET [DK/DK]; Anker Engelunds Vej 101 A, 2800 Kgs. Lyngby (DK).

(72) Inventors: SIMONSEN, Henrik Toft; Mellemvang 14B, 2970 Horsholm (DK). PERAMUNA, Anantha Vithakshana; Rentemestervej 24, 2.th., 2400 Copenhagen NV (DK).

(74) Agent: PATENTGRUPPEN A/S; Aaboulevarden 31, 4th floor, DK-8000 Aarhus C (DK).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available):* AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,

(54) Title: PRODUCTION OF TERPENES AND TERPENOIDS

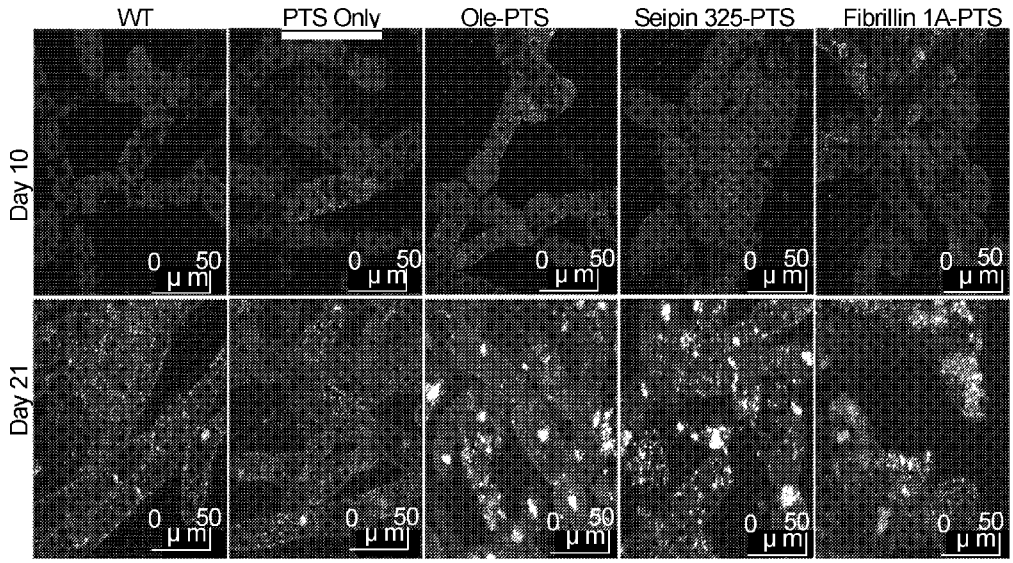

Fig. 2

(57) Abstract: This invention relates to a transgenic bryophytic cell capable of producing geranyl pyrophosphate (GPP), farnesyl-pyrophosphate (FPP) and/or geranylgeranyl pyrophosphate (GGPP) comprising at least one heterologous nucleic acid molecule encoding at least one chimeric protein comprising at least a first and a second polypeptide where said first and second polypeptides are operationally linked and said first polypeptide is a lipid body-associated protein. The invention further relates to a method for producing the transgenic bryophytic cell as well as a method for preparing terpenes and terpenoids using the transgenic bryophytic cells.

WO 2019/057258 A1

[Continued on next page]

TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

PRODUCTION OF TERPENES AND TERPENOIDS

FIELD OF THE INVENTION

The present invention relates to a transgenic bryophytic cell capable of producing terpenes and terpenoids, and methods of producing the terpenes and terpenoids as well as methods of producing the transgenic bryophytic cell.

BACKGROUND OF THE INVENTION

Terpenoids or terpenes are natural products found in a multiplicity of organisms (bacteria, fungi, animals, plants). The compounds consist of isoprene units (C5H8) and are classified by the number of units present in their structure. Thus, monoterpenes, sesquiterpenes and diterpenes are terpenes containing 10, 15 and 20 carbon atoms respectively.

The common five-carbon precursor to all terpenes is isopentenyl pyrophosphate (IPP). IPP forms the acyclic prenyl pyrophosphate terpene precursors for each class of terpenes, e.g. geranyl pyrophosphate (GPP) for the monoterpenes, farnesyl-pyrophosphate (FPP) for the sesquiterpenes, and geranylgeranyl-pyrophosphate (GGPP) for the diterpenes. These precursors serve as substrate for the terpene synthases or cyclases, which are specific for each subclass of terpene, e.g. monoterpene, sesquiterpene or diterpene synthases. Some terpene synthases produce a single product, but most of them produce multiple products. The synthases are responsible for the extremely large number of terpene skeletons. Finally, in the last stage of terpenoid biosynthesis, the terpene molecules may undergo several steps of secondary enzymatic transformations such as hydroxylations, isomerisations, oxido-reductions or acylations, leading to the tens of thousands of different terpene molecules.

Monoterpenes, sesquiterpenes and diterpenes accumulates in plants and can be extracted by different means such as steam distillation or solvent extraction that produces the so-called essential oil containing the concentrated terpenes. Such natural plant extracts are important components for the flavor and perfume industry due to their flavor and fragrance properties, and some monoterpenes, sesquiterpenes and diterpenes may even possess cosmetic, medicinal and antimicrobial effects. Extracted *terpens* molecules are often used as such, but in some cases chemical reactions are used to transform the terpenes into even higher valued molecules.

Because **of** the complexity **of** the terpene structure, production **of** individual terpene molecules by chemical synthesis is often limited by the cost **of** the process and may not always be chemically **or** financially feasible. The price and availability **of** the plant natural extracts is dependent **on** the abundance, the oil yield and the geographical origin **of** the plants. It has recently been shown that moss *Physcomitrella patens (P. patens)* is able to produce terpenoids like Patchoulol, sclareol and β-santalene when transduced with heterologous nucleic acid encoding relevant synthases for the production **of** these sesquiterpenes and diterpene (WO 2014/206412).

However, the terpenes and terpenoids as produced in WO 2014/206412 are heterologous products to the bryophytic cells and high amounts of terpenes and terpenoids is likely to be detrimental to the cells and be able to influence the production. Thus, it would be difficult to obtain an efficient production of the terpenes and terpenoids. Furthermore, it is difficult to produce complex terpenes and terpenoids, which may require more biosynthetic steps to be created.

OBJECT OF THE INVENTION

It is the object of the invention to develop bryophytic cells which are capable of producing terpenes and terpenoids in high amounts, without the production being detrimental to the cell.

It is a further object of the invention to develop bryophytic cells which are capable of producing complex terpenes and terpenoids.

SUMMARY OF THE INVENTION

It has surprisingly been found in this invention that a transgenic bryophytic cell capable of producing geranyl pyrophosphate (GPP), farnesyl-pyrophosphate (FPP) and/or geranylgeranylpyrophosphate (GGPP) comprising at least one heterologous nucleic acid molecule encoding at least one chimeric protein comprising at least a first and a second polypeptide where said first and second polypeptides are operationally linked and said first polypeptide is a lipid body-associated protein and said second polypeptide is a biosynthetic enzyme, is able to direct the production of terpenes and terpenoids to the oil bodies of the bryophytic cells. Hereby, the produced terpenes and terpenoids are not accumulated in the cytoplasm of the cells and raising to an increased level, which may be detrimental to the cells.

In contrast, the terpenes and terpenoids are localized in the oil bodies and do not obstruct with the growth of the cells. The retainment of produced terpenes and terpenoids in the oil bodies is thus increased compared to cells not comprising the lipid body-associated protein. Retaining terpene and terpenoid volatile compounds, such as e.g. Patchoulol, in the lipid bodies may also improve the isolated yield of the compounds because they otherwise would evaporate during production. Furthermore, the localization of the terpenes and terpenoids to the oil bodies enables the terpenes and terpenoids to be easily purified from the bryophytic cells. Thus, The inventor surprisingly show that it is possible to use lipid bodies to produce and retain other compounds such as terpenes and terpenoids that are not related to lipid metabolism. In addition, by linking the chimeric proteins to the oil bodies, more biosynthetic enzymes may be closely related to one another by them being arranged in close proximity instead of being distributed in the entire cell. Hereby, the product from the first biosynthetic enzyme is readily passed on the next biosynthetic enzyme for which it is a substrate. Usually, kinetics (efficacy) of enzymes are influenced by physical linking them, which may negatively influence the production by for example poor kinetics. Surprisingly, the inventors of the present invention have found that they are able to direct production of compounds to lipid bodies in the cell and increase the retainment in the lipid bodies when a lipid body-associated protein and a biosynthetic enzyme is operationally linked.

In one embodiment, said lipid body-associated protein is selected from the group of oleosin, fibrillin, seipin, perilipin, small rubber particle protein 1, small rubber particle protein 2 and small rubber particle protein 3.

In a further embodiment, said biosynthetic enzyme is selected from the group of monoterpene synthases, sesquiterpene synthases, cytochrome P450, alcohol dehydrogenase, acetyl transferase, aldehyde reductase, aldehyde dehydrogenase.

In one embodiment, said at least one chimeric protein encodes a third polypeptide being operationally linked to said first or said second polypeptide.

Hereby, the second and third polypeptide may be in close proximity of one another. This is particularly advantageous when the second and third polypeptides are different biosynthetic enzymes making up consecutive steps in a biosynthetic pathway.

In one embodiment, said at least said first and second polypeptides are separated by a linker.

By separating the polypeptides by a linker the genes are separated by an inert gene sequence and the three-dimensionally structuring of the separate proteins may be correctly obtained.

Thus, the functioning of the proteins, which is often influenced by correct three-dimensionally folding, may be achieved more efficiently.

In a further embodiment, said linker is a flexible linker such as a GS-linker or a rigid linker.

In one embodiment, said cell comprises at least a first and a second heterologous nucleic acid molecule, said first heterologous nucleic acid encoding a first chimeric protein and said second heterologous nucleic acid molecule encoding a second chimeric protein, where both said first chimeric protein and said second chimeric protein encode at least a first polypeptide being a lipid body-associated protein and a second polypeptide being a biosynthetic enzyme.

Transducing the cell with at least a first and a second heterologous nucleic acid molecule enables different second polypeptides to be expressed in the cell. Furthermore, the second polypeptides will be linked to the lipid bodies and is likely in close proximity of one another. This could be highly advantageous especially for obtaining complex products where that could be the result of a multistep procedure.

This invention also includes a bryophyte plant or bryophytic tissue comprising transgenic bryophytic cells as described herein.

This invention also describes a method of producing a transgenic bryophytic cell comprising introducing into a bryophytic cell capable of producing geranylpyrophosphate (GPP), farnesyl-pyrophosphate (FPP) and/or geranylgeranylpyrophosphate (GGPP) at least one heterologous nucleic acid molecule encoding at least one chimeric protein comprising at least a first and a second polypeptide where said first and second polypeptides are operationally linked and said first polypeptide is a lipid body-associated protein and said second polypeptide is a biosynthetic enzyme.

This invention also describes a method for preparing terpenes and terpenoids in a transgenic bryophytic cell comprising the steps of:

a) Introducing into a bryophytic cell capable of producing geranyl pyrophosphate (GPP), farnesyl-pyrophosphate (FPP) and/or geranylgeranyl pyrophosphate (GGPP) at least one heterologous nucleic acid molecule encoding at least one chimeric protein comprising at least a first and a second polypeptide where said first and second polypeptides are operationally linked and said first polypeptide is a lipid body-associated protein and said second polypeptide is a biosynthetic enzyme,

b) culturing the transgenic bryophytic cell to express or overexpress said at least one chimeric protein, hereby enabling the transgenic bryophytic cell to produce terpenes and terpenoids,

c) isolating the terpenes or terpenoids produced from the transgenic bryophytic cell.

In one embodiment, step c) is performed by isolating the oil bodies of the transgenic bryophytic cell prior to isolating the terpenes or terpenoids.

In a further embodiment, the method for preparing terpenes and terpenoids relates to said biosynthetic enzyme being a monoterpene synthase or a sesquiterpene synthase.

This invention also describes the use of a bryophytic cell as described herein and/or a bryophyte plant as described herein for production terpenes and terpenoids.

DESCRIPTION OF THE DRAWINGS

Figure 1. illustrates detection of Patchoulol as measured by GC-MS. From the top down Patchoulol standard, PTS only, PTS-Seipin, PTS-SRP-1, PTS-Oleosin, PTS-Fibrillin.

Figure 2. illustrates transgenic bryophytic cells transduced with different heterologous nucleic acids having a first polypeptide being a lipid body associated protein and a second polypeptide being Venus fluorescence protein.

Figure 3. Map of pRH004 including the positions for the "Insert Check" primer pair. See also SEQ ID No 22, SEQ ID No 23 and SEQ ID No 26. pRH004 is an artificial DNA sequence from designed for biological research.

Figure 4. Map of plasmid containing the Venus fluorescence gene. See also SEQ ID No 27. This plasmid is an artificial DNA sequence from designed for biological research.

Figure 5. Illustrates protein co-localization to lipid droplets

Figure 6. Illustrates what is trapped inside the cells including lipid droplets

Figure 7. Illustrates the presence of patchoulol inside the lipid droplets. Oil bodies were extracted from PpOlel- **GSio** -PTS line.

Figure 8. Illustrates a higher image resolution of a verification of patchoulol inside the lipid droplets.

DETAILED DESCRIPTION OF THE INVENTION

Abbreviation and terms:

The term "bryophyte cell" and "bryophytic cell" is used interchangeably and is to be understood cells derived from mosses, hornworts and liverworts.

The term "moss cells" is to be understood as bryophytic cells belonging to the class Bryophyta and its subclasses.

The term "WT" or "wild-type" is to be understood as a bryophytic cell, which does not comprise the heterologous nucleic acid according to the invention but is a non-transduced bryophytic cell.

By the term "heterologous nucleic acid molecule" is to be understood a nucleic acid molecule which is introduced into the bryophytic cell in order for it to be expressed in the bryophytic cell. The heterologous nucleic acid molecule may encode one or more proteins such as fusion proteins or chimeric proteins. One or more of the proteins may be unknown to the bryophytic cell.

The term "chimeric protein" is to be understood as a fusion polypeptide i.e. a polypeptide where two or more genes originally coding for separate proteins are fused. Translation of the fusion gene results in a chimeric protein with functional properties derived from each of the original polypeptides. The phrases "operatively positioned", "operatively linked", "under control" and "under transcriptional control" mean that the separate proteins are in a correct functional location and/or orientation in relation to one another to control transcription of functional proteins.

The term "promoter" is to be understood as commonly known to the person skilled in the art i.e. a promoter is a control sequence. The promoter is typically a region of a nucleic acid sequence at which initiation and rate of transcription are controlled. It may contain genetic elements at which regulatory proteins and molecules may bind such as RNA polymerase and other transcription factors. The phrases "operatively positioned", "operatively linked", "under control" and "under transcriptional control" mean that a promoter is in a correct functional location and/or orientation in relation to a nucleic acid sequence to control transcriptional initiation and expression of that sequence.

The term "terminator" is to be understood as commonly known to the person skilled in the art i.e. a terminator is a control sequence. The terminator is typically a region of a nucleic acid

sequence at which the transcription is stopped. The phrases "operatively positioned", "operatively linked", "under control" and "under transcriptional control" mean that a terminator is in a correct functional location and/or orientation in relation to a nucleic acid sequence to control transcriptional termination of that sequence.

By the "term" linker is to be understood an amino acid or peptide-mimetic sequence, which may be inserted between proteins.

The term "primer" is to be understood as to encompass any nucleic acid that is capable of priming the synthesis of a nucleic acid in a template-dependent process.

The term "protein", as used herein, includes proteins, polypeptides, and peptides.

The terms "protein", "amino acid sequence" and "polypeptide" are used interchangeably herein. In the present disclosure and claims, the conventional one-letter and three-letter codes for amino acid residues may be used. The 3-letter code for amino acids as defined in conformity with the IUPACIUB Joint Commission on Biochemical Nomenclautre (JCBN). It is also understood that a polypeptide may be coded for by more than one nucleotide sequence due to the degeneracy of the genetic code. In addition, it is to be understood that persons skilled in the art may, using routine techniques, make nucleotide substitutions that do not affect the protein/polypeptide/amino acid sequence encoded by the nucleotide sequence of the invention to reflect the codon usage of any particular bryophytic cell in which the protein according to the present invention is to be expressed.

The terms "variant" or "derivative" in relation to the nucleic acid sequences of the present invention includes any substitution of, variation of, modification of, replacement of, deletion of or addition of one (or more) nucleic acids from or to the sequence providing the resultant amino acid sequence has the same functionality or activity as the parent amino acid sequence.

The terms "transformed cell", "transduced cell" or "transgenic cell" are used interchangeable and include cells that have been transformed by use of recombinant DNA techniques. The transformation typically occurs by insertion of one or more nucleotide sequences into the cell to be transformed.

The term "biosynthetic enzyme" as used herein, means any enzyme involved in biosynthesis also known as biogenesis or anabolism. In one embodiment of this invention biosynthetic enzyme relates to enzymes involved in the conversion of compounds to terpenes and terpenoids.

By the term, "terpene synthase", is meant, unless specifically limited, a polypeptide capable of catalyzing the synthesis of a monoterpene, sesquiterpene or a diterpene in the form of any of its stereoisomers or a mixture thereof, starting from GPP, FPP or GGPP, respectively.

By the term, "sesquiterpene synthase", is meant, unless specifically limited, a polypeptide capable of catalyzing the synthesis of a sesquiterpene in the form of any of its stereoisomers or a mixture thereof, starting from FPP.

By the term, "monoterpene synthase", is meant, unless specifically limited, a polypeptide capable of catalyzing the synthesis of a monoterpene in the form of any of its stereoisomers or a mixture thereof, starting from GPP.

The term "lipid body associated protein" or "oil body associated protein" as used herein means a polypeptide being associated with the oil bodies in the plants i.e. be integrated in the membrane of the oil bodies or being connected with other proteins or components of the membrane of the oil body unless specifically limited.

The terms "lipid body" and "oil body" are used interchangeable herein.

Bryophytic cell

Bryophyte is the traditional name used to refer to all embryophytes (land plants) that are non-vascular plants such as mosses, hornworts and liverworts. The bryophytes differ considerably from the vascular plants in several aspects. The major differences between bryophytes and vascular plants are the lack of the specialized structures to carry water and nutrients i.e. bryophytes do not contain xylem and phloem. The water and nutrients must be dispersed via diffusion to the tissues of bryophytes. Thus, most bryophytes lack complex tissue-organization and are relatively small compared with most seed-bearing plants. The bryophytes also vary in terms of their life stages and dominant forms from higher, vascular plants where the bryophytes are characterized by a dominant gametophytic stage. The bryophytes are widely distributed throughout the world.

Hornwort (division Anthocerotophyta) is also called horned liverwort and comprises about 600 species of small non-vascular plants. They normally grow on damp soils or on rocks in tropical and warm temperate regions. They are subdivided into five classes; Leiosporocerotaceae, Anthocerotaceae, Notothyladaceae, Phymatocerotaceae and Dendrocerotaceae. Liverworts (division: Marchantiophyta) is subdivided into three classes; Haplomitriopsida, Marchantiopsida and Jungermanniopsida.

Liverwort (division Marchantiophyta) are small non-vascular spore-producing plants. They comprise more than 9.000 species. Liverworts are distributed worldwide but most commonly in the tropics.

Mosses (division Bryophyta) are small non-vascular spore bearing land plants. They comprise at least 14.000 species. The mosses are distributed throughout the world except in salt water and are commonly found in moist shady locations. They are subdivided into eight classes; Takakiopsida, Sphagnopsida, Andreaeopsida, Andreaeobryopsida, Oedipodiopsida, Polytrichopsida, Tetraphidopsidaan and Bryopsida,

In one embodiment, the bryophytic cells are moss cells.

In one embodiment of the present invention, the bryophytic cells belong to the class of Bryopsida.

In one embodiment, the bryophytic cell is a moss cell.

In a further embodiment, the moss cell is selected from the group comprising Takakiopsida, Sphagnopsida, Andreaeopsida, Andreaeobryopsida, Oedipodiopsida, Polytrichopsida, Tetraphidopsidaan and Bryopsida.

In a further embodiment, the bryophytic cells belong to the genus Physcomitrella such as but not limited to *Physcomitrella patens {P.patens).*

In a still further embodiment, said moss cell is *Physcomitrella patens.*

Physcomitrella contains all the cellular compartments relevant for terpene and terpenoid biosynthesis found in higher plants, such as the endoplacmic reticulum (ER) and plastids. In addition, codon usage is conserved between Physcomitrella and higher plants such as Arabidopsis thaliana, as are several posttranslational modifications such a N-glycosylationa (Rensing SA *et al,* 2005); Vietor R *et al,* 2003; Koprivova A *et al,* 2003; Mega T, 2007). Thus, Physcomitrella has strong potential to heterologously express functional enzymes with same functionality as the endogenous higher plant (Simonsen HT *et al.,* 2009). This was demonstrated by successful expression of Patchoulol, β-santalene and sclareol in Physcomitrella (WO20 14/2064 12). Physcomitrella represents an ancient lineage of land plants, and its metabolic and chemical diversity is low compared to higher plants. This is illustrated by the number of cytochromes P450 (P450s) and UDP glycosyltransferased (UGTs) found in the genome. The genomes of Arabidopsis thaliana and Oryza sativa contain 246 and 343 P450s respectively, while the genome of Physcomitrella only contains 71 P450s

(Hamberger B *et al,* 2013). Similarly, the genome of Phycomitrella contains a low number of UGTs compared to other land plants (Yonekura-Sakakibara K *et al.,* 201 1). The low number of P450s and UGTs found in Physcomitrella and the correspondingly lower chemical diversity reduces the risk of unspecific modifications by endogenous enzymes, through pathways used in higher plants for detoxification of xenobiotics. In addition to this Physcomitrella has a simple terpenoid profile and the geneome of Physcomitrella only contains a single functional terpene synthase (TPS) (Chen F *et al.,* 2011). Gene editing by efficient homologous recombination in Physcomitrella provides a very powerful tool for metabolic engineering (Schaefer DG *etal.,* 1997). Knocking out moss terpenoid synthesis genes can be accomplished by homologous recombination see e.g. (Bach SS *et al.* 2013). In addition to having a terpenoid free-background, GPP, FPP or GGPP (the universal precursors for monoterpene, sesquiterpene and diterpene biosynthesis) could be redirected into heterologous expressed terpenoid pathways.

Processes for culturing moss cells are known in the art (Decker EL *et al.* 2008; Knight *et al.,* 2002; Cove DJ *et al.,* 2009). Similar methods can be used for the culturing of liverworts and hornworts. Thus, the culturing of a transgenic bryophytic cell described herein can be carried out in accordance with such processes.

Transgenic bryophytic cells/plants

There are several methods known in the art for the creation of transgenic plants and also in moss cells. These are well described in the art (King, Brian Christopher, *et al.,* 2016; Bach SS *etal,* 2014; Reski R, 1999; Reski R, Botanica Acta 1998; Sugano *etal,* 2014; Lopez-Obando *etal,* 2016; Chiyoda *etal,* 2008; Kubota *et al,* 2013).

Protoplast transformation is the most commonly used method for moss transformation including Physcomitrella transformation, and is well described in the literature (King, Brian Christopher, *et al,* 2016; Bach SS *et al,* 2014). The method requires careful handling and regeneration of fragile protoplasts and must be done under sterile conditions. This method can be very efficient and yield a large number of stable transformants, however, several attempts may be needed to successfully recover stable moss lines. Another robust alternative to e.g. PEG mediated transformation of protoplasts is biolistic transformation, i.e. direct gene transfer by particle bombardment may be utilized. In another embodiment, agrobacterium-mediated transformation may be utilized.

In one embodiment, the transformation may be introduced into the cytosol. In another embodiment, the transformation may be introduced into the chloroplasts.

Direct gene transfer by particle bombardment provides an example for transforming plant tissue. In this technique a particle, or micro projectile, coated with DNA is shot through the physical barrier of the cell. Particle bombardment can be used to introduce DNA into any target tissue that is penetable by DNA coated particles, but for stable transformation, it is imperative that regenerable cells be used. Typically, the particles are made of gold or tungsten. The particles are coated with DNA using either CaCh or ethanol precipitation methods which are commonly known in the art. DNA coated particles are shot out of a particle gun. A suitable particle gun can be purchased from Bio-Rad Laboratoires (Hercules, Calif). Particle penetration is controlled by varying parameters such as the intensity of the explosive burst, the size of the particles, or the distance particles must travel to reach the target tissue. The DNA used for coating the particles may comprise an expression cassette suitable for driving the expression of the gene of interest that will comprise a promoter operably linked to the gene of interest. For example, moss transformation protocols are described in (Reutter, K *et al.,* 1996; Reski, R, Plant Biology, 1998).

In one embodiment, the bryophytic cells may be transformed using a delivery vehicle comprising a nucleic acid molecule according to any embodiment of the invention. The person skilled in the art is capable of selecting a suitable delivery vehicle.

In one embodiment the delivery vehicle includes the nucleic acid molecule of the invention operably linked to at least one regulatory sequence which controls transcription, translation, initiation and termination such as a transcriptional promoter, operator or enhancer, or an mRNA ribosomal binding site and, optionally, including at least one selection marker.

In one embodiment of the invention, the transgenic bryophytic cells containing a heterologous nucleic acid as described herein may be identified *in vitro* or *in vivo* by encoding a screenable or selectable marker in the heterologous nucleic acid.

When transcribed and translated, a marker confers an identifiable change to the bryophytic cell permitting identification of bryophytic cells containing the heterologous and/or homologous nucleic acid molecule.

A positive selectable marker is one which the presence of the marker allows for its selection, while a negative selectable marker is one in which the presence of the marker prevents its selection.

In one embodiment, the selection marker is selected from the group of Kanamycin, Hygromycin, Sulfadizine sodium and Gentamicin sulfate.

Biosynthetic enzyme

A heterologous nucleic acid molecule encoding a biosynthetic enzyme such as terpene synthase or a combination of several heterologous nucleic acid molecules encoding biosynthetic enzymes such as terpene synthases as the second polypeptide may be transduced into a bryophytic cell. A heterologous nucleic acid molecule encoding the biosynthetic enzyme may be isolated from any suitable organism e.g. prokaryotes, plant cells or eukaryotes.

In a further embodiment, said biosynthetic enzyme is selected from the group of monoterpene synthases, sesquiterpene synthases, cytochrome P450, alcohol dehydrogenase, acetyl transferase, aldehyde reductase, aldehyde dehydrogenase.

In a still further embodiment, said biosynthetic enzyme is a monoterpene synthase or a sesquiterpene synthase.

In one embodiment, said second polypeptide is a biosynthetic enzyme selected from the group of β-pinene synthase (EC 4.2.3.120), limonene synthase (EC 4.2.3.114), γ-terpinene synthase (EC 4.2.3.114), S-linalool synthase (EC 4.2.3.25), 1,8-cineol synthase (EC 4.2.3.108), bornyl diphosphate synthase (EC 5.5.1.8), sabinene synthase (EC 4.2.3.110), ocimene synthase (EC 4.2.3.106), fenchol synthase (EC 4.2.3.10), borneol synthase, cadinene synthase (EC 4.2.3.92), Patchoulol synthase (EC 4.2.3.70), valencene synthase (EC 4.2.3.73), bisabolol synthase, α-humulene synthase (EC 4.2.3.104), CYP71BA1 (EC 1.14.13.150), zerumbone synthase (EC 1.1.1.326), santalene synthase (EC 4.2.3.82), CYP736A167, geraniol synthase (EC 3.1.7.11), geraniol dehydrogenase (EC 1.1.1.183), geraniol acetyltransferase.

In a still further embodiment, the amino acid sequence of the polypeptide having monoterpene synthase activity has at least 80% sequence identity to an amino acid sequence selected from among SEQ ID No 29, 31, 33, 35, 37, 39, 41, 43, 45 and 47.

In a still further embodiment, the amino acid sequence of the polypeptide having sesquiterpene synthase activity has at least 80% sequence identity to an amino acid sequence selected from among SEQ ID No 47, 49, 51 and 53.

In a further embodiment, said transgenic bryophytic cell comprises three different second polypeptide encoded by three different chimeric proteins where said three different second

polypeptides are α-humulene synthase (EC 4.2.3.104), CYP71BA1 (EC 1.14.13.150) and zerumbone synthase (EC 1.1.1.326)

In a further embodiment, said transgenic bryophytic cell comprises two different second polypeptide encoded by two different chimeric proteins where said two different second polypeptides are santalene synthase (EC 4.2.3.82) and CYP736A167.

In a further embodiment, said transgenic bryophytic cell comprises three different second polypeptide encoded by three different chimeric proteins where said three different second polypeptides are geraniol synthase (EC 3.1.7.11), geraniol dehydrogenase (EC 1.1.1.183) and geraniol acetyltransferase.

Examples of biosynthetic enzymes are described in table 1.

Group	Enzyme	EC number	Ref. Number	Product	CAS No
Single protein; Monoter-pene synthases	(-)-β-pinene synthase	EC 4.2.3.120	AF5 14288	β-pinene	127-91-3
	(+)-limonene synthase	EC 4.2.3.20	AF5 14287	limonene	5989-27-5
	γ-terpinene synthase	EC 4.2.3.114	AF5 14286	γ-terpinene	99-85-4
	S-linalool synthase	EC 4.2.3.25	KF700700	linalool	78-70-6
	1,8-cineol synthase	EC 4.2.3.108	AF05 1899	1,8-cineol	470-82-6
	(+)-bornyl di¬ phosphate synthase	EC 5.5.1.8	AF05 1900	bornyl diphosphat e	64822-87-3
	sabinene synthase	EC 4.2.3.110	AF05 1901	sabinene	3387-41-5
	ocimene synthase	EC 4.2.3.106	AY575970	β -ocimene	13877-91-3
	fenchol synthase	EC 4.2.3.10	AY693648	fenchol	1632-73-1
	borneol synthase		AB 120957	borneol	507-70-0
Single protein; Sesqui¬ terpene synthases	cadinene synthase	EC 4.2.3.92	NM_00 1297453	γ -cadinene	39029-41-9
	patchoulol synthase	EC 4.2.3.70	KP694233	patchoulol	5986-55-0
	valencene synthase	EC 4.2.3.73	NM_00 1288856	valencene	4630-07-3
	bisabolol synthase		KM259907	bisabolol	23089-26-1
Multiple proteins	1) a-humulene synthase 2) CYP71BA1 3) zerumbone synthase	1) EC 4.2.3.104 2) EC 1.14.13.150 3) EC 1.1.1.326	AB24733 1 AB33 1234 AB48083 1	zerumbone	471-05-6
Multiple proteins	1) Santalene synthase 2) CYP736A167	1) EC 4.2.3.82	HQ343276 KU169302	a -santalol	115-71-9
Multiple proteins	1) geraniol synthase 2) geraniol dehydrogenase 3) geraniol acetyltransferase	1) EC 3.1.7.11 2) EC 1.1.1.183	AB69153 1 AY879284 AY850287	geranyl acetate	105-87-3

Table 1: Biosynthetic enzymes

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be β-pinene synthase (EC 4.2.3.120) that converts geranyl diphosphate (GPP) to β-pinene (CAS no 127-91-3) and diphosphate. In one specific embodiment, the β-pinene synthases has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of β-pinene synthase derived from citrus limon (SEQ ID No. 29). In a further specific embodiment, the β-pinene synthases has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of β-pinene synthase derived from citrus limon (SEQ ID No. 28).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be limonene synthase (EC 4.2.3.20) that converts geranyl diphosphate (GPP) to (4R)-limonene (CAS no 5989-27-5) and diphosphate. In one specific embodiment, the limonene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of limonene synthase derived from *Citrus limon* (SEQ ID No. 31). In a further specific embodiment, the limonene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of limonene synthase derived from *Citrus limon* (SEQ ID No. 30).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be γ-terpinene synthase (EC 4.2.3.114) that converts geranyl diphosphate (GPP) to γ-terpinene (CAS no 99-85-4) and diphosphate. In one specific embodiment, the γ-terpinene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of γ-terpinene synthase derived from *Citrus limon* (SEQ ID No. 33). In a further specific embodiment, the γ-terpinene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of γ-terpinene synthase derived from *Citrus limon* (SEQ ID No. 32).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be S-linalool synthase (EC 4.2.3.25) that converts geranyl diphosphate (GPP) and water to linalool (CAS no 78-70-6) and diphosphate. In one specific embodiment, the S-linalool synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of S-linalool synthase derived from *Coriandrum sativum* (SEQ ID No. 35). In a further specific embodiment, the S-linalool synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of S-linalool synthase derived from *Coriandrum sativum* (SEQ ID No. 34).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be 1,8-cineol synthase (EC 4.2.3.108) that converts geranyl diphosphate (GPP) and water to 1,8-cineole (CAS no 470-82-6) and diphosphate. In one specific embodiment, the 1,8-cineol synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of 1,8-cineol synthase derived from *Salvia officinalis* (SEQ ID No. 37). In a further specific embodiment, the 1,8-cineol synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of 1,8-cineol synthase derived from *Salvia officinalis* (SEQ ID No. 36).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be bornyl diphosphate synthase (EC 5.5.1.8) that converts geranyl diphosphate (GPP) to bornyl diphosphate (CAS no 64822-87-3). In one specific embodiment, the bornyl diphosphate synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of bornyl diphosphate synthase derived from *Salvia officinalis* (SEQ ID No. 39). In a further specific embodiment, the bornyl diphosphate synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of bornyl diphosphate synthase derived from *Salvia officinalis* (SEQ ID No. 38).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be sabinene synthase (EC 4.2.3. 110) that converts geranyl diphosphate (GPP) to sabinene (CAS no 3387-41-5) and diphosphate. In one specific embodiment, the sabinene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of sabinene synthase derived from *Salvia officinalis* (SEQ ID No. 41). In a further specific embodiment, the sabinene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of sabinene synthase derived from *Salvia officinalis* (SEQ ID No. 40).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be ocimene synthase (EC 4.2.3.106) that converts geranyl diphosphate (GPP) to β-ocimene (CAS NO 13877-91-3) and diphosphate. In one specific embodiment, the ocimene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of E-beta-ocimene synthase derived from *Lotus corniculatus var. japonicus* (SEQ ID No. 43). In a further specific embodiment, the ocimene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of E-beta-ocimene synthase derived from *Lotus corniculatus var. japonicus* (SEQ ID No. 43).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be fenchol synthase (EC 4.2.3.10) that converts geranyl diphosphate (GPP) and water to fenchol (CAS no 1632-73-1) and diphosphate. In one specific embodiment, the fenchol synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of fenchol synthase derived from *Ocimum basilicum* (SEQ ID No. 45). In a further specific embodiment, the fenchol synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of fenchol synthase derived from *Ocimum basilicum* (SEQ ID No. 45).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be borneol synthase (EC N/A) that converts geranyl diphosphate (GPP) to borneol (CAS no 507-70-0) and diphosphate. In one specific embodiment, the borneol synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of limonene/borneol synthase derived from *Chamaecyparis obtusa col* (SEQ ID No. 47). In a further specific embodiment, the borneol synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of limonene/borneol synthase derived from *Chamaecyparis obtusa col* (SEQ ID No. 46).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be γ-cadinene synthase (EC 4.2.3.92) that converts farnesyl diphosphate (FPP) to γ-cardinene (CAS no 39029-41-9) and diphosphate. In one specific embodiment, the γ-cadinene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of γ-cadinene synthase derived from *Cucumis melo* (SEQ ID No. 49). In a further specific embodiment, the γ-cadinene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of γ-cadinene synthase derived from *Cucumis melo* (SEQ ID No. 48).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be Patchoulol synthase (EC 4.2.3.70) that converts farnesyl diphosphate (FPP) and water to Patchoulol (CAS no 5986-55-0) and diphosphate. In one specific embodiment, the Patchoulol synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of Patchoulol synthase derived from *Pogostemon cablin* (SEQ ID No. 51). In a further specific embodiment, the Patchoulol synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of Patchoulol synthase derived from *Pogostemon cablin* (SEQ ID No. 50).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be valencene synthase (EC 4.2.3.73) that converts farnesyl diphosphate (FPP) to valencene (CAS no 4630-07-3) and diphosphate. In one specific embodiment, the valencene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of valencene synthase derived from *Citrus sinensis* (SEQ ID No. 53). In one specific embodiment, the valencene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of valencene synthase derived from *Citrus sinensis* (SEQ ID No. 52).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be bisabolol synthase (EC N/A) that converts farnesyl diphosphate (FPP) to bisabolol (CAS no 23089-26-1) and diphosphate. In one specific embodiment, the bisabolol synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of a-bisabolol synthase derived from *Matricari chamomilla var. recutita* (SEQ ID No. 55). In a further specific embodiment, the bisabolol synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of a-bisabolol synthase derived from *Matricari chamomilla var. recutita* (SEQ ID No. 54).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be **ot**-humulene synthase (EC 4.2.3.104) that converts farnesyl diphosphate (FPP) to α-humulene (CAS no 6753-98-6) and diphosphate. In one specific embodiment, the **ot**-humulene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of **ot**-humulene synthase derived from *Zingiber zerumbet zssl* (SEQ ID No. 57). In a further specific embodiment, the **ot**-humulene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of **ot**-humulene synthase derived from *Zingiber zerumbet zssl* (SEQ ID No. 56).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be CYP71BA1 (EC 1.14.13.150) also known as **ot**-humulene 8-hydroxylase or **ot**-humulene 8-hydroxylase that converts **ot**-humulene to 8-hydroxy-**ot**-humulene (synonym 10-hydroxy-a-humulene, zerumbol) (17678-70-5). In one specific embodiment, the CYP71BA1 has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of P450 mono-oxygenase derived from *Zingiber zerumbet* (SEQ ID No. 59). In a further specific embodiment, the CYP71BA1 has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence

identity to the nucleic acid sequence of P450 mono-oxygenase derived from *Zingiber zerumbet* (SEQ ID No. 58).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be zerumbone synthase (EC 1.1.1.326) that converts 10-hydroxy-a-humulene to zerumbone (CAS no 471-05-6). In one specific embodiment, the zerumbone synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of the short-chain dehydrogenase/reductae 1 derived from *Zingiber zerumbet zsdl* (SEQ ID No. 61). In a further specific embodiment, the zerumbone synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of the short-chain dehydrogenase/reductae 1 derived from *Zingiber zerumbet zsdl* (SEQ ID No. 60).

In a further embodiment, the biosynthetic enzymes α-humulene synthase, CYP71BA1 and zerumbone synthase are expressed in the same bryophytic cell converting FPP to zerumbone. In a still further embodiment, the biosynthetic enzymes are expressed on different chimeric proteins. In an even further embodiment, at least two of the enzymes are expressed in the same chimeric protein. In one embodiment, α-humulene synthase and CYP71BA1 are expressed on the same chimeric protein. In a further embodiment, CYP71BA1 and zerumbone synthase are expressed on the same chimeric protein. In a further embodiment, **ot**-humulene synthase, CYP71BA1 and zerumbone synthase are all expressed by the same chimeric protein.

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be **ot**-santalene synthase (EC 4.2.3.82) that converts farnesyl diphosphate (FPP) to a-santalene (CAS no 512-61-8) and diphosphate. In one specific embodiment, the a-santalene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of santalene synthase derived from *Santalum album* (SEQ ID No. 63). In a further specific embodiment, the a-santalene synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of santalene synthase derived from *Santalum album* (SEQ ID No. 62).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be CYP736A167 (EC N/A) that converts a-santalene to a-santalol (CAS no 115-71-9). In one specific embodiment, the CYP736A167 has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of CYP736A167 derived from *Santalum album* (SEQ ID No. 65). In a further specific embodiment, the CYP736A167 has at least 70, 75, 80,

85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of CYP736A167 derived from *Santalum album* (SEQ ID No. 64).

In a further embodiment, the biosynthetic enzymes α-santalene synthase and CYP736A167 are expressed in the same bryophytic cell converting FPP to a-santalol. In a still further embodiment, the biosynthetic enzymes are expressed on different chimeric proteins. In an even further embodiment, at two biosynthetic enzymes are expressed by the same chimeric protein.

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be geraniol synthase (EC 3.1.7.11) that converts geranyl diphosphate (GPP) to geraniol (CAS no 106-24-1) and diphosphate. In one specific embodiment, the geraniol synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of geraniol synthase derived from *Citrusjambhiri RlemTPS3* (SEQ ID No. 67). In a further specific embodiment, the geraniol synthase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of geraniol synthase derived from *Citrusjambhiri RlemTPS3* (SEQ ID No. 66).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be geraniol dehydrogenase (EC 1.1.1.183) that converts geraniol to geranial (CAS no 96839-90-6). In one specific embodiment, the geraniol dehydrogenase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of geraniol dehydrogenase (GEDH) derived from *Ocimum basilicum* (SEQ ID No. 69). In a further specific embodiment, the geraniol dehydrogenase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of geraniol dehydrogenase (GEDH) derived from *Ocimum basilicum* (SEQ ID No. 68).

In one embodiment, the biosynthetic enzyme expressed as a chimeric protein may be geraniol acetyltransierase (EC N/A) that converts geranial to geranyl acetate (CAS no. 105-87-3). In one specific embodiment, the geraniol acetyltransferase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of acetyl CoA geraniol/citronellol acetyltransferase (AATl) derived from *Rosa hybrid cultivar* (SEQ ID No. 71). In a further specific embodiment, the geraniol acetyltransferase has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of acetyl CoA geraniol/citronellol acetyltransferase (AATl) derived from *Rosa hybrid cultivar* (SEQ ID No. 70).

In a further embodiment, the biosynthetic enzymes geraniol synthase, geraniol dehydrogenase and geraniol acetyltransferase are expressed in the same bryophytic cell converting GPP to

geranyl acetate. In a still further embodiment, the biosynthetic enzymes are expressed on different chimeric proteins. In an even further embodiment, at least two of the enzymes are expressed in the same chimeric protein. In one embodiment, geraniol synthase and geraniol dehydrogenase are expressed by the same chimeric protein. In a further embodiment, geraniol dehydrogenase and geraniol acetyltransferase are expressed by the same chimeric protein. In a further embodiment, geraniol synthase, geraniol dehydrogenase and geraniol acetyltransferase are all expressed by the same chimeric protein.

Lipid body associated protein

Most plant cells comprise one or more oil bodies, which is a lipid-containing structure. The term oil body may refer to at least two distinct kinds of structures in different kind of plants. In liverworts, the oil bodies are structures unique to liverworts which contain isoprenoid essential oils and are surrounded by a single membrane. The size, shape, color and number of oil bodies per cell is characteristic of certain species and may be used to identify these. Plant TAGs are mainly stored in the small organelles called oil bodies that are assembled in different tissues in the plant (Murphy, DJ, 2005). The oil bodies are an organelle that has evolved to hold triglycerides in the plant cells. Thus, they are the principal store of chemical energy. It seems that proteins out-number lipids on the surface of the oil bodies with a domination of the molecule oleosin. The lipid and protein fractions of the oil bodies are remarkable because they maintain a coherent monolayer over a wide temperature and hydration range.

The first polypeptide of the chimeric protein is a lipid body associated protein, which includes oil body forming proteins as well as other proteins being in close relation with the lipid body. Oleosin, oleosin-like or oil body forming proteins (OBFP) are required to form the membrane of the oil body. Oil body forming proteins are proteins like oleosin, seipins and small rubber proteins. These proteins normally comprise a central hydrophobic domain that is proposed to make up a structure that interact with the lipids, and enable the monolayer membrane to form.

Lipid body associated proteins which are able to associate itself with the membrane of the lipid body/bodies in the bryophytic cell may be used in the chimeric protein to be transduced into the bryophytic cell. None limiting examples hereof are oleosin, fibrillin, seipin, perilipin, small rubber particle protein 1, small rubber particle protein 2 and small rubber particle protein 3. Further non-limiting examples are described in table 2.

Lipid Body-associated Protein	For example (Ref. Number)
Oleosin	**PpOlel** (XM_00 1766345.1)
Seipin	PpSeipinlS325 **(Pp3c8_820Vl.l)** (www.cossmoss.org)
Seipin	At Seipin 2 (NM_102716.4)
Small Rubber Protein 1	At SRPl (NM_179525.3)
Fibrillin	At Fibrillin 1A (NM_1 16640.5)
Fibrillin	At Fibrillin **IB** (NM_1 18350.3)

Table 2: lipid body associated protein

In one embodiment, the lipid body associated protein expressed as the first polypeptide of the chimeric protein according to the invention may be oleosin. In one specific embodiment, the oleosin has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of the predicted protein (PHYPADRAFT_80169) derived from *Physcomitrella patens subsp. patens* (SEQ ID No. 73). In a further specific embodiment, the oleosin has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of the predicted protein (PHYPADRAFT_80169) derived from *Physcomitrella patens subsp. patens* (SEQ ID No. 72).

In one embodiment, the lipid body associated protein expressed as the first polypeptide of the chimeric protein according to the invention may be seipin. In one specific embodiment, the seipin has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of the Pp3c8_820Vl.l derived from *Physcomitrella patens* (SEQ ID No. 75). In a further specific embodiment, the seipin has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of the Pp3c8_820Vl.l derived from *Physcomitrella patens* (SEQ ID No. 74). In one specific embodiment, the seipin has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of the putative adipose-regulatory protein (seipin) derived from *Arabidopsis thaliana* (SEQ ID No. 77). In a further specific embodiment, the seipin has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of the putative adipose-regulatory protein (seipin) derived from *Arabidopsis thaliana* (SEQ ID No. 76).

In one embodiment, the lipid body associated protein expressed as the first polypeptide of the chimeric protein according to the invention may be Small Rubber Protein 1. In one specific embodiment, the Small Rubber Protein 1 has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of the Rubber elongation factor protein (REF)

derived from *Arabidopsis thaliana* (SEQ ID No. 79). In a further specific embodiment, the Small Rubber Protein 1 has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of the Rubber elongation factor protein (REF) derived from *Arabidopsis thaliana* (SEQ ID No. 78).

In one embodiment, the lipid body associated protein expressed as the first polypeptide of the chimeric protein according to the invention may be Fibrillin 1A. In one specific embodiment, the Fibrillin 1A has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of the fibrillin (FIB) derived from *Arabidopsis thaliana* (SEQ ID No. 81). In a further specific embodiment, the Fibrillin 1A has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of the fibrillin (FIB) derived from *Arabidopsis thaliana* (SEQ ID No. 80).

In one embodiment, the lipid body associated protein expressed as the first polypeptide of the chimeric protein according to the invention may be Fibrillin IB. In one specific embodiment, the Fibrillin IB has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the amino acid sequence of the Plastid-lipid associated protein PAP/fibrillin family protein derived from *Arabidopsis thaliana* (SEQ ID No. 83). In a further specific embodiment, the Fibrillin IB has at least 70, 75, 80, 85, 90, 95, 98, 99 or 100 % sequence identity to the nucleic acid sequence of the Plastid-lipid associated protein PAP/fibrillin family protein derived from *Arabidopsis thaliana* (SEQ ID No. 82).

Chimeric protein

In one embodiment, said cell comprises at least a first and a second heterologous nucleic acid molecule, said first heterologous nucleic acid encoding a first chimeric protein and said second heterologous nucleic acid molecule encoding a second chimeric protein, where both said first chimeric protein and said second chimeric protein encode at least a first polypeptide being a lipid body-associated protein and a second polypeptide being a biosynthetic enzyme.

If more than one heterologous nucleic acid encoding for at least one chimeric protein is introduced into the bryophytic cell, the first polypeptide of the chimeric proteins may be different or similar. As an example, if two chimeric proteins are encoded in the bryophytic cell the first polypeptide of the first chimeric protein may be oleosin and the first polypeptide of the second chimeric protein may be oleosin. As another example, if three chimeric proteins are encoded in the bryophytic cell the first polypeptide of the first chimeric protein may be

oleosin, the first polypeptide of the second chimeric protein may be oleosin and the first polypeptide of the third chimeric protein may be seipin. As a further example if three chimeric proteins are encoded in the bryophytic cell, the first polypeptide of the first chimeric protein may be seipin, the first polypeptide of the second chimeric protein may be fibrillin 1A and the first polypeptide of the third chimeric protein may be Small Rubber Protein 1.

In a further embodiment, the second polypeptide of said first and said second chimeric protein are different polypeptides.

By introducing more than one heterologous nucleic acid into the bryophytic cell a higher amount of the second polypeptide may be obtained if the second polypeptides are similar. If the second polypeptides are different it will be possible to link different polypeptides to the oil bodies and hereby arrange the second polypeptides in close proximity. This may be beneficial particularly if the second polypeptides each forms part of a synthetic pathway i.e. the product of one polypeptide could be the substrate for the next polypeptide.

The second polypeptide may be a biosynthetic enzyme and is operationally linked to the first polypeptide. If more than one heterologous nucleic acid encoding for at least one chimeric protein is introduced into the bryophytic cell, the second polypeptide of the chimeric proteins may be different or similar. In a still further embodiment, said second polypeptides of said first chimeric protein and said second chimeric protein are different biosynthetic enzymes.

The chimeric protein may further comprise a third polypeptide, which may be similar to or different from the second polypeptide. In a further embodiment, said third polypeptide is a biosynthetic enzyme. In a further embodiment, said third polypeptide is different from said second polypeptide.

The chimeric protein may further comprise a fourth polypeptide, which may be similar to or different from the second and/or third polypeptide. The chimeric protein may further comprise a fifth polypeptide, which may be similar to or different from the second, third and/or fourth polypeptide. The chimeric protein may further comprise a sixth polypeptide, which may be similar to or different from the second, third, fourth and/or fifth polypeptide.

Introducing even further second polypeptides into the transgenic bryophyte cell enables complex biosynthetic pathways to be formed i.e. complex final molecules such as terpenes and terpenoids may be formed.

According to this invention several combinations of first and second polypeptides in the chimeric protein is a possibility. In some embodiments, the combinations of the first and the second polypeptide in the chimeric protein are as described in table 3.

First polypeptide	Second polypeptide	Optional linker
Oleosin	Patchoulol synthase	Flexible
Seipin	Patchoulol synthase	Flexible
Fibrillin 1A	Patchoulol synthase	Flexible
Fibrillin 1B	Patchoulol synthase	Flexible
Oleosin	Patchoulol synthase	Rigid
Oleosin	Patchoulol synthase	Cleavable

Table 3: Embodiments of combinations

Linkers

In the chimeric protein the first polypeptide and the second polypeptide may be separated by a linker. This may advantageously introduce better folding of both the first and second polypeptide. Furthermore, the function of the first and second polypeptide may be more easily preserved by obtaining a correct folding of the polypeptide, which is not influenced by steric hindrance of the other polypeptide to which it is coupled. If further polypeptides are encoded in the chimeric protein linkers may as well be arranged between these further polypeptides and their neighboring polypeptides.

Linkers may have one or more properties that may include a flexible conformation, an inability to form an ordered secondary structure or a hydrophobic or charged character, which could promote or interact with either domain.

In one embodiment of this invention, the linker is a flexible linker. Amino acids typically found in flexible protein regions include Gly, Asn and Ser. In a specific embodiment, the flexible linker is a glycine-serine linker (GS-linker). In a further specific embodiment, the flexible linker is as described in SEQ ID No 84.

In one embodiment of this invention, the linker is a rigid linker to prohibit unwanted interactions. In a specific embodiment, the rigid linker is a (EAAAK)3 linker. In a further specific embodiment, the rigid linker is as described in SEQ ID No 85.

None-limiting examples of linker sequences which may be used in this invention is described in table 4.

Protein-Protein Linker	Sequence
Flexible linker (glycine-serine)10	5'-GGTAGCGGCAGCGGTAGCGGTAGCGGCAGC-3'
Rigid linker (EAAAK)3	5'-GAAGCTGCTGCTAAGGAAGCTGCTGCTAAGGAAGCTGCTGCTAAG-3'
Transcriptionally cleaved linker (LP4/2A)	5'-TCAAATGCAGCAGACGAAGTTGCTACTCAACTTTTGAATTTTGACTTGCTGAAGTTGGCTGGTGATGTTGAGTCAAACCCTGGACCT-3'

Table 4: linker sequences

Primers

Paris of primers designed to selectively hybridize to nucleic acids corresponding to sequences of the proteins of interest are contacted with the template nucleic acid under conditions that permit selective hybridization. Depending upon the desired application, high stringency hybridization conditions may be selected that will only allow hybridization to sequences that are completely complementary to the primers. In other embodiments, hybridization may occur under reduced stringency to allow for amplification of nucleic acids containing one or more mismatches with the primer sequences. Once hybridized, the template-primer complex is contacted with one or more enzymes that facilitate template-dependent nucleic acid synthesis. Multiple rounds of amplifications, also referred to as "cycles" are conducted until a sufficient amount of amplification product is produced.

Typically, primers are oligonucleotides from ten to twenty and/or thirty base pairs in length, but longer sequences can be employed. Primers may be provided in double-stranded and/or single stranded form, although the single-stranded form is preferred.

Amplified product

The amplified product may be detected or quantified. In certain applications, the detection may be performed by visual means as commonly known by the person skilled in the art (Gallagher, Sean R. *etal.*, 2007).

The encoding of the heterologous nucleic acid molecule in the bryophytic cell is arranged by operably linking the heterologous nucleic acid molecule to a promoter. According to the present invention, the promoter may be any promoter functional in a bryophytic cell. As an example a Ubiquitin promoter from Maize or an Actin promoter from rice.

The particular promoter that is employed to control the expression of the protein encoding nucleic acid molecule of the invention is not believed to be critical as long as it is capable of expressing the nucleic acid molecule in the bryophytic cell.

Furthermore, the encoding of the heterologous nucleic acid molecule in the bryophytic cell is arranged operably linking the heterologous nucleic acid molecule to at least one terminator or termination signal. These are comprised of DNA sequences involved in specific termination of an RNA transcript by an RNA polymerase. Terminators contemplated for use in the invention include any known terminator of transcription described herein or known to one of ordinary skill in the art, including but not limited to, for example a NOS terminator from rice or an OCS terminator as long as they are functional in a bryophytic cell.

Amino acid sequence/nucleic acid sequence identity

The present invention covers variants, homologues or derivatives of the amino acid sequences presented herein, as well as variants, homologues or derivatives of the nucleotide sequence coding for those amino acid sequences.

The percentage of identity between two nucleic acid sequences/amino acid sequences is a function of the number of nucleic acid residues/amino acids that are identical in the two sequences when an alignment of these two sequences has been generated. Identical residues are defined as residues that are the same in the two sequences in a given position of the alignment. The percentage of sequence identity, as used herein, is calculated from the optimal alignment by taking the number of residues identical between two sequences dividing it by the total number of residues in the shortest sequence and multiplying by 100. The optimal alignment is the alignment in which the percentage of identity is the highest possible. Gaps

may be introduced into one or both sequences in one or more positions of the alignment to obtain the optimal alignment. These gaps are then taken into account as non-identical residues for the calculation of the percentage of sequence identity.

Alignment for the purpose of determining the percentage of nucleic acid sequence identity/amino acid sequence identity can be achieved in various ways using computer programs such as publicly available computer programs. Preferably, the BLAST program set to be default parameters, available from the National Center for Biotechnology Information (NCBI) at http://www.ncbi.nlm.nih.gov/BLAST/bl2seq7wblast2.cgi, can be used to obtain an optimal alignment of amino acid sequences and nucleic acid sequences and to calculate the percentage of sequence identity.

Generation of variant nucleotides having the above required percent identities is common general knowledge. Thus, a person skilled in the art could generate a large number of nucleotide and/or polypeptide variants having, for example, at least 70%-99% identity to the reference sequence described herein and screen such for desired phenotypes according to methods routine in the art e.g. in (King, Brian Christopher, *et al.,* 2016).

Nucleic acids used as a template for amplification may be isolated from cells according to standard methodologies known to the person skilled in the art. The nucleic acid may be genomic DNA, fractionated or whole cell RNA. When using RNA, it may be desired to first convert the RNA to a complementary DNA by standard techniques commonly known to the persons skilled in the art.

All publications, patents, patent applications, and other references cited in this application are incorporated herein by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent allocation or other reference was specifically and individually indicated to be incorporated by reference in its entirety for all purposes. Citation of a reference herein shall not be construed as an admission that such is prior art to the present invention.

EXAMPLES

The following examples illustrate various embodiments of the present invention and provides techniques for establishing Physcomitrella as a transgenic bryophytic cell for stable production of hydrophobic compounds.

Each of the individual examples should not be construed as limiting the scope of the present invention.

Plant material, growth conditions and transformation

Wild type *P. patens* (Gransden ecotype) was obtained from the International Moss Stock Center at the University of Freiburg (http://www.moss-stock-center.org/). Plasmid material for Arabidopsis was purchased from Arabidopsis Stock Centre (https://www.arabidopsis.Org/index.j sp)

Growth conditions and transformation processes are similar to those described in detail in (Bach SS *etal,* 2013).

Growth media for Physcomitrella patens

1. Phy B media, modified from minimal media: For 1L mix 800 mg $Ca(N03)_2$, 250 mg $MgS0_4$-$7H_20$, 12.5 mg $FeS0_4$-$7H_20$, 10 mL KH_2P0_4 buffer (25 g KH_2P0_4 per liter and adjusted to pH 6.5 with 4M KOH) and 0.25 mL trace element solution (110 mg $CuS0_4$-$5H_20$, 110 mg $ZnS0_4$-$7H_20$, 1228 mg H_3B0_3, 778 mg $MnCI_2$-$4H_20$, 110 mg $CoCI_2$-$6H_20$, 53 mg Kl, 50 mg Na_2Mo0_4-$2H_20$ per liter). The minimal medium was supplemented with 0.5 g ammonium tartrate per liter. The medium can be solidified with 0.8 % (w/v) Agar A, and was sterilized by autoclaving at 121 °C.

2. BCD media: For one-liter mix 12.5 mg $FeS0_4 \cdot 7H_20$, 10 mL solution B (25 g/L $MgS0_4$-$7H_20$), 10 mL solution C (25 g/L KH_2P0_4 adjusted to pH 6.5 with 4 M KOH), 10 mL solution D (101 g/L KNO3), 1 mL trace element solution (614 mg H3BO3, 389 mg $MnCI_2$-$4H_20$, 55 mg $AI_2(S0_4)_3$ K_2S0_4-$24H_20$, 55 mg $CoCI_2$-$6H_20$, 55 mg $CuS0_4$-$5H_20$, 55 mg $ZnS0_4$-$7H_20$, 28 mg KBr, 28 mg Kl, 28 mg LiCI, 28 mg and $SnCI_2$-$2H_20$ per liter). The medium was solidified with 0.8 % (w/v) Agar A and was sterilized by autoclaving at 121°C. After autoclaving, 10 mL sterile 1 M $CaCI_2$ solution was added (10 mM final concentration).

Table 5 provides an overview of the constructs used.

Lipid Body-associated Protein	**Name (species)**	**NCBI entry**	**SEQ ID No**
Oleosin	PpOle1 *{Physcomitrelia patens)*	XM_001766345. 1	72-73
Seipin	PpSeipinlS325 *{Physcomitrelia patens)*	Pp3c8_820Vl. 1	74-75
Seipin	At Seipin 2 *{Arabidopsis thaliana)*	NM_102716.4	76-77
Small Rubber Protein 1	At SRP1 *{Arabidopsis thaliana)*	NM_179525.3	78-79
Fibrillin	At Fibrillin 1A *{Arabidopsis thaliana)*	NM_l 16640.5	80-81
Fibrillin	At Fibrillin IB *{Arabidopsis thaliana)*	NM_1 18350.3	82-83
Linkers			
Flexible	GS-linker		84
Rigid	(EAAAK)3		85
Transcriptionally cleaved	LP4/2A		86
Second polypeptide			
Flourescence protein	Venus		27
Patchoulol synthase	PTS		50-5 1

Table 5: Overview of constructs

Table 6 provides an overview of the primers used.

5

Name	F/R*	Sequence
4.7 kb	F	5'- CCAGATCGACCACATCCTTCTCCG-3' (SEQ ID No 1)
4.7 kb	R	5'-GACCTGCAGAAGTAACACCAAACAACAG-3' (SEQ ID No 2)
2.1 kb	F	5'- GTCCTGCTTTAATGAGATATGCGAGACG-3' (SEQ ID No 3)
2.1 kb	R	5'- ACGAAGGCCGTTCTTCCCTG-3' (SEQ ID No 4)
GSio-Venus	F	5'- GGTAGCGGCAGCGGTAGCGGTAGCGGCAGC-ATGGTGAGCAAGGGCGAGG-3'(SEQ ID No 5)
AtSeipin2	F	5'- CTGTTGTTTGGTGTTACTTCTGCAGGTCATGGACTCCGAGTCCGAG-3'(SEQID No 6)
GSio-AtSeipin2	R	5'- GGTAGCGGCAGCGGTAGCGGTAGCGGCAGC-CCTCCAGACTCCAGTTGG-3'(SEQ ID No 7)
AtSRPl	F	5'-CTGTTGTTTGGTGTTACTTCTGCAGGTCATGGCTGAAGATGAAATAG-3' (SEQ ID No 8)
GSio-AtSRPl	R	5'- GGTAGCGGCAGCGGTAGCGGTAGCGGCAGC-ATCAGCTCGACACTGATC-3'(SEQ ID No 9)
ppOlel	F	5'-CTGTTGTTTGGTGTTACTTCTGCAGGTCATGGATAATGCCAAAACCAAGGC-3' (SEQ ID No 10)
GSio-ppOlel	R	5'- GGTAGCGGCAGCGGTAGCGGTAGCGGCAGC-AGCCGCGACGCTGGTATC-3'(SEQ ID No 11)
PP1S325	F	5'-CTGTTGTTTGGTGTTACTTCTGCAGGTCATGGCTTCCTCCGACGTC-3' (SEQ ID No 12)
GSio-PP1S325	R	5'- GGTAGCGGCAGCGGTAGCGGTAGCGGCAGC-GTTTTGATCCAGAACCTTTCC-3'(SEQ ID No 13)
FiblA	F	5'-CTGTTGTTTGGTGTTACTTCTGCAGGTCATGGCGACGGTACCATT-3' (SEQ ID No 14)
GSio-FiblA	R	5'- GGTAGCGGCAGCGGTAGCGGTAGCGGCAGC-AGGGTTTAAGAGAGAGCTTCC-3'(SEQID No 15)
FibIB	F	5'-CTGTTGTTTGGTGTTACTTCTGCAGGTCATGGCGACGGTACAATT-3'(SEQ ID No 16)
GSio-FiblB	R	5'- GGTAGCGGCAGCGGTAGCGGTAGCGGCAGC-AGGATTCAAGAGAGGG-3'(SEQ ID No 17)
GSio-PTS	F	5'- GGTAGCGGCAGCGGTAGCGGTAGCGGCAGC-ATGGAGTTGTATGCCCAAAGT-3' (SEQ ID No 18)
PTS	R	5'-CGTCTCGCATATCTCATTAAAGCAGGACTTAATATGGAACAGGGTGAAGG-3' (SEQ ID No 19)
PTS only	F	5'-CTGTTGTTTGGTGTTACTTCTGCAGGTCATGGAGTTGTATGCCCAAAG-3'(SEQ ID No 20)
PTS only	R	5'-CGTCTCGCATATCTCATTAAAGCAGGACTTAATATGGAACAGGGTGAAG-3 (SEQ ID No 21)

Table 6: Overview of primers. F-forward primer, R-reverse primer (continues on next page).

Name	F/R*	Sequence
Insert Check	F	5'-CCTGCCTTCATACGCTATTTATTTGCT-3'(SEQ ID No 22)
Insert Check	R	5'-CAACGTGCACAACAGAATTGAAAGC-3'(SEQ ID No 23)
EAAAK3-ppOlel	R	5'-GAAGCTGCTGCTAAGGAAGCTGCTGCTAAGGAAGCTGCTGCTAAG-AGCCGCGACGCTGGTATC-3' (SEQ ID No 24)
EAAAK3-PTS	F	5'-GAAGCTGCTGCTAAGGAAGCTGCTGCTAAGGAAGCTGCTGCTAAG-ATGGAGTTGTATGCCCAAAGT-3' (SEQ ID No 25)
GS-linker		5'- GGTAGCGGCAGCGGTAGCGGTAGCGGCAGC-3'(SEQ ID No 84)
Rigid-linker		5'-GAAGCTGCTGCTAAGGAAGCTGCTGCTAAGGAAGCTGCTGCTAAG-3'(SEQ ID No 85)
LP4/2A linker		5'-TCAAATGCAGCAGACGAAGTTGCTACTCAACTTTTGAATTTTGACTTGCTGAAGTTGGCTGGTGATGTTGAGTCAAACCCTGGACCT-3' (SEQ ID No 86)

Table 6 continued: Overview of primers. F-forward primer, R-reverse primer.

EXAMPLE 1

Preparation of cDNA from *P. patens*

1μg of total RNA extracted from 7-day-after-blending *Physcomitrella* cells was used for cDNA synthesis using iScript™ cDNA synthesis kit (Bio-rad, 170-8891), by following manufacturer's instruction.

Preparation of cDNA from *A. thaiiana*

1μg of total RNA extracted from A.thaliana seeds were used for cDNA synthesis using iScript™ cDNA synthesis kit (Bio-rad, 170-8891), by following manufacturer's instruction.

Preparation of PpOlel, PpSeipinlS325, AtSeipin 2, At Small rubber protein 2, AtFibrillin 1A and AtFibrillin 1B for lipid droplet (LD) or endoplasmic reticulum (ER) co-localization

DNA fragment containing PpOlel, PpSeipinlS325, AtSeipin 2, AT Small rubber protein 1 (SRP1), AtFibrillin 1A and AtFibrillin lB_genes with flexible linker to be assembled *in vivo* with Venus fluorescence protein, which is a mutated and improved version of yellow

fluorescence protein. Following the fragments were integrated into the *Physcomitrella* genome.

The DNA fragments were constructed as follows:

1) 4.7 kb region with 108 5' neutral locus, G418 selection marker with CaMV 35S promoter/ CaMV poly(A) signal and Maize Ubiquitin promoter was amplified from the pRH004 plasmid (SEQ ID NO 26) by PCR using the primer pair 4.7 kb (Table 6; SEQ ID No 1 and SEQ ID No 2).

2) Venus fluorescent protein sequence was amplified along with OCS terminator and 108 3' neutral locus from the pRH004 venus plasmid (SEQ ID No 27) by using the primer pair sequence (Table 6; SEQ ID No 4 and SEQ ID No 5). See Figure 4.

3) PPOlel, PPSeipinlS325 gene coding sequences was amplified from *P. patens* cDNA by PCR and AtFibrillin 1A and FibrillinlB, AtSeipin 2, At small rubber protein 2 gene coding sequence was amplified from *A. thaliana* cDNA with over-hang primers to 4.7 kb and over-hang with flexible, rigid or LP4/2A sequence (Table 6; SEQ ID No 6-17).

Preparation of PpOlel, PpSeipinlS325, AtFibrillin 1A and AtFibrillin IB fused Patchoulol synthase (PTS) with flexible, rigid or LP42A linker for transformation

DNA fragment containing PPOlel, PPSeipinlS325, Fibrillin 1A and Fibrillin IB genes with flexible, rigid or LP42A linker to be assembled *in vivo* with PTS and integrated into the *Physcomitrella* genome were constructed as follows:

1) 4.7 kb region with 108 5' neutral locus, G418 selection marker with CaMV 35S promoter/ CaMV poly(A) signal and Maize Ubiquitin promoter was amplified from the pRH004 plasmid (SEQ ID No 26) by PCR using the primer pair 4.7 kb (Table 6; SEQ ID No 1-2).

2) 2.1 kb region with the OCS terminator and 108 3' neutral locus was amplified from the pRH004 plasmid by PCR using the prime pair 2.1 kb (Table 6; SEQ ID No 3-4).

3) PpOlel, PpSeipinlS325 gene coding sequences was amplified from *P. patens* cDNA by PCR and AtFibrillin 1A and AtFibrillinlB gene coding sequence was amplified from *A. thaliana* cDNA with over-hang primers to 4.7 kb and over-hang with flexible or rigid sequence (Table 6; see if rigid liSEQ ID No 10-17 and 24).

4) PTS gene coding sequence was amplified by PCR with over-hang primers to flexible or rigid and over-hang with 2.1 kb region (Table I; SEQ ID No 18-19 and 25).

5) The DNA fragments were purified with QIAquick PCR Purification Kit from Qiagen before used for transformation

Preparation of PTS only fragment for transformation

This fragment does not link PTS with a lipid body associated protein but only PTS is expressed by the fragment and is used as a control.

1) 4.7 kb region with 108 5' neutral locus, G418 selection marker with CaMV 35S promoter/ CaMV poly(A) signal and Maize Ubiquitin promoter was amplified from the pRH004 plasmid (SEQ ID No 26) by PCR using the primer pair 4.7 kb (Table I; SEQ ID No 1-2).

2) 2.1 kb region with the OCS terminator and 108 3' neutral locus was amplified from the pRH004 plasmid by PCR using the prime pair 2.1 kb (Table I; SEQ ID No 3-4).

3) PTS gene coding sequence was amplified by PCR with over-hang primers to 4.7 kb and over-hang primers to 2.1 kb region (Table I; SEQ ID No 20-21).

4) The DNA fragments were purified with QIAquick PCR Purification Kit from Qiagen before used for transformation

EXAMPLE 2

Preparation and transformation of protoplasts

1) On the day of transformation, new PEG solution was prepared, and allowed to stand for 2 hours before use. The solution was subsequently sterilized by filtration by passing it through a 0.22 µm syringe filter.

2) 5-7 day old *P.patens* protonema tissue was harvested by scraping it off cellophane-overlaid agar plates and placed in a sterile 50 mL plastic tube. 1 mL of a 0.5% Driselase solution was prepared for every 40 mg tissue. The Driselase powder was dissolved in 8.5% D-mannitol, and sterilized using a syringe filter and Physcomitrella tissue was added subsequently. The mixture was incubated for 30-60 minutes with occasional inversion of tube until the tissue had been thoroughly digested.

3) The Driselase-treated tissue was poured through a sterile 100 μm stainless steel mesh screen, and the protoplasts were recovered in a sterile beaker. Undigested tissue and cellular debris do not pass through the mesh.

4) The protoplasts were centrifuged at 200 rcf for 5 minutes, with gentle breaking.

5) The supernatant was decanted using a serological pipette.

6) The pellet was re-suspended in protoplast wash solution using the same volume as driselase in step 2.

7) The steps 4 and 5 were repeated.

8) The protoplasts were re-suspended in half the original volume of 8.5% D-mannitol and the density was estimated using a hemocytometer.

9) The solution was centrifuged at 200 rcf for 5 minutes, and the supernatant removed. The pellet was re-suspended in sterile MMM solution yielding a protoplast concentration of 1.5-2 x 10^6 protoplasts/mL.

10) 30 μL (approximately 5-10 μg) of PCR amplified DNA was added to the bottom of a 15 mL round bottom tube. 300 μL of protoplast suspension and 300 μL of sterile PEG solution were added and mixed with the DNA by flicking the tube.

11) The mixture was incubated at 45°C in a water bath for 5 minutes followed by 5 minutes at room temperature.

12) The protoplast suspension was diluted 4 times with 300 μL of 8.5% D-mannitol, followed by an additional 4 times dilution with 1 mL of 8.5% D-mannitol.

13) The transformed protoplasts were centrifuged at 200 rcf with gentle braking for 5 minutes, and the supernatant was removed.

14) The protoplasts were resuspended in 500 μL 8.5% D-mannitol, and 2.5 mL of molten PRMT was added.

15) 1 mL of the protoplast suspension was added to a PRMB petri dish overlaid with sterile cellophane. At least 3 plates were made from each transformation event.

16) The plates were sealed with 3M micropore tape, and placed in a growth chamber under standard conditions.

17) The protoplasts were allowed to regenerate their cell walls for 5-7 days, and then the cellophane and regenerating plants were proceeded to selective media and selection of positive transformants as described in Example 3.

EXAMPLE 3

Selection procedure to obtain stable transgenic bryophytic cells.

1) Cellophane discs with transformed moss were transferred onto solid PhyB media with 30 µg/mL G418 or 30 µg/mL hygromycin and incubated for two weeks under standard conditions.

2) The cellophane discs were transferred with recovered transformants to solid PhyB media and incubated for another 2 weeks for relaxation. Unstable or transient transformants loose the plasmid and the ability to survive on selective media in this period.

3) Step 1 and 2 was repeated. Positive moss lines obtained after two rounds of selection were considered stably transformed bryophytic cells.

4) The stably transformed bryophytic cells were checked for correct insertion of the nucleotide sequence using the "Insert Check" forward and reverse primers (Table 6, SEQ ID No 22-23).

EXAMPLE 4 - Protein co-localization to protein droplets

DNA fragment containing PpOlel, PpSeipinlS325, AtSeipin 2, AT Small rubber protein 1 (SRP1), AtFibrillin 1A and AtFibrillin IB genes were tagged with Venus fluorescent protein and a flexible linker were prepared as described in Example 1.

Materials & Methods

Staining for lipid droplets

1) Suspend cells in suspended in PBS pH 7.4 buffer.

2) Stain the cells with lipophilic dye BODIPY 505/515 (Thermofisher,D3921) with a final concentration of 3µg/mL BIODIPY.

3) Incubate the cells in the dark for 10 minutes.

4) Visualize the lipid droplets with a 488 nm laser excitation line and a 510-530 nm emission window. Positively stained lipid droplets will emit a green fluorescence.

5) Visualize the chloroplast using the same laser line and 650-700nm emission window. Chloroplasts will emit a red fluorescence.

6) Perform a Z- stack with a line average of 4.

7) Combine the Z-stacks to a single image.

Venus fluorescent protein co-localization of lipid droplet (LD)/endoplasmic reticulum proteins (ER)

1) Stain and visualize the cells for LDs and chloroplasts as described above.

2) Visualize the localization of Venus fluorescent protein with 514 nm laser excitation line and 535-550 nm emission window. The Venus fluorescent protein will emit a yellow fluorescence.

3) Perform a Z-stack and combine the image as above.

Results

Staining for lipid droplets

10 and 21 days old PpOlel-(gly-ser)10-PTS, PpSeipinlS325-(gly-ser)10-PTS, and AtFibrillin 1B-(gly-ser)10-PTS, PTS only and wild type cells were stained for lipid droplets with BIODIPY 505/515. All lines but WT had considerably higher amount of lipid droplets with PpOlel-(gly-ser)10-PTS, PpSeipinlS325-(gly-ser)10-PTS having the most amount of lipid droplets.

Figure 2 illustrates transgenic bryophytic cells transduced with different heterologous nucleic acids and table 7 provides a quantitative assessment of the results.

Table 7: Quantitative assessment of the amount of lipid droplets

Line	Linker	10 days	21 days
WT	None	+	++
PpOle1	Flexible (GS_{10})	++++	+++++
PpOle1	Rigid (EAAAK)3	++++	+++++
PpSeipin1S325	flexible	++++	+++++
AtFibrillin 1A	flexible	++	+++++
PTS	none	++	++++

Protein co-localization to lipid droplets

PpOle 1 protein localizes to LDs and ER with a higher concentration of the protein localizing to LDs. Both PpSeipinlS325 and AtSeipin 2 concentrates specifically to ER where the biogenesis of LDs are taking place. On the other hand, SRPl localizes specifically to lipid droplets while AtFibrillins localizes into globular structures in chloroplast as illustrated in figure 5.

EXAMPLE 5 - Detection of Patchoulol

DNA fragment containing PpOlel, PpSeipinlS325, AtSeipin 2, AT Small rubber protein 1 (SRPl), AtFibrillin 1A and AtFibrillin IB genes were tagged with PTS with different linkers.

Materials & Methods

Cell preparation

1) Cell lines were prepared for analysis of Patchoulol by growing the strains along with the wild type cells

2) The medium was blended and changed every week for three weeks.

Detection

1) Presence of Patchoulol was detected using a previously published method (Zhan X *et al.*, 2014).

Quantification of Patchoulol

1) Grow cells in 20 ml culture flasks for more than a week.

2) Filtered the cells to remove media.

3) Extract the media with hexane for any additional Patchoulol in the media.

4) Add 500µl of di water to the cells with 500µE of ethyl acetate and lyse the cells using a bead beater and place the cells in the ultrasonic bath for 30 minutes.

5) Extract the upper ethyl acetate layer and transfer the extract to a new vial and analyze by GC-MS.

6) Measure the dry weight of the extracted cells and calculate the amount of Patchoulol per gram of cells

Patchoulol composition in LD

1) Make protoplasts as mentioned above.

2) Centrifuge the protoplast and remove any additional liquid buffer.

3) Add 10ml of di water to the pellet to lyse the cells.

4) Add sucrose to the lysate to a final concentration of 0.3M sucrose.

5) Load the mixture to an ultracentrifuge tube with a 2 ml over-lay of water.

6) Centrifuge for 60 minutes, 135,000 x g, room temperature in a swinging bucket roter.

7) Extract the top water layer with the lipid droplets.

8) Repeat step 4-7.

9) Extract the content of the lipid droplets with 100% ethyl acetate (V/V).

10) Transfer the extract to a new vial and analyze by GC-MS.

10) Measure the dry weight of the protoplast and calculate the amount of Patchoulol per gram of protoplast

11) Compare the composition of Patchoulol to other compounds in LDs

Results

Detection of Patchoulol

Detection of Patchoulol is illustrated in figure 1. Patchoulol was detected from PpOlel- GSio -PTS, PpSeipinlS325- GSio -PTS, AtSRPl-GSio-PTS, and AtFibrillin 1A- GSio -PTS lines. From the top down in Figure 1 is illustrated Patchoulol standard, PTS only, PpSeipinlS325- GSio -PTS, AtSRPl-GSio-PTS, PpOlel- GSio -PTS and AtFibrillin 1A- GSio -PTS. The detection in PTS only line had the highest concentration of Patchoulol inside the cells followed by PpOlel- GSio -PTS, PpSeipinlS325- GSio -PTS, AtFibrillin 1A- GSio -PTS, and PpOlel-(EAAAK)3-PTS lines. Upon, normalizing for the expression of Patchoulol synthase, PpSeipinlS325- GSio -PTS gave the highest theoretical yield of Patchoulol. So when normalized by gene expression, theoretically *Seipin325-PTS* will produce more Patchoulol. AtFibrillin 1A yielded the lowest. The precursor needed for Patchoulol production, Farnesyl pyrophosphate (FPP), is present at a very low concentration in chloroplast compared to cytoplasm of the cell. Thus, the production of Patchoulol is low in the chloroplast without the plastid targeted over-expression of FPP. Overall, the lipid droplet targeted Patchoulol synthase had a lower expression comparted to Patchoulol synthase over expressed by itself. By expressing extra copies of lipid droplet targeted Patchoulol synthase the production of Patchoulol may be largely increased.

Quantification of patchoulol

The data of figure 6 is showing what is trapped inside the cells including lipid droplets. Figure 6a is showing Patchoulol inside cells and figure 6b is showing Patchoulol inside cells. When Patchoulol is normalized to the expression of PTS S235-PTS (seipin) gave the highest yield. The absolute amount of Patchoulol retained in the lipid bodies can be calculated as (retention rate)*(total Patchoulol inside the cell).

Presence of Patchoulol in LD

Oil bodies were extracted from PpOlel- **GSio** -PTS line and the presence of patchoulol inside the lipid droplets was verified as can be seen from figure 7. A higher image resolution of a verification of patchoulol inside the lipid droplets is shown in figure 8. The upper part is a chromatographic trace and MS of the peak at 15.47 is shown in the lower left part of the figure. The peak at 222 corresponds to Patchoulol.

Table 8 is showing production inside the cell and percentage retained in the cell in the lipid bodies. The percentage of retention is (Patchoulol from isolated lipid body)/(Patchoulol inside the cell)* 100. The results in table 8 shows that all but OLE-LP4/2A-PTS retain more Patchoulol in the lipid bodies compared to PTS alone. LP4/2A is a protease cleavage site and as a result OLE and PTS become separate proteins. This may be used as a control because it shows that when the enzymes are not attached physically, the retention rate is lower.

Cell line	Production (µg/g)	Ret %
PTS	33	31
OLE-LP4/2A-PTS	17	27
OLE-PTS	4.5	43
SRP-PTS	13	41
Seipin-PTS	5	37

Table 8. Production of Patchoulol inside the cell and percentage Patchoulol retained in the cell in the lipid bodies.

SEQUENCES

SEQ **ID** NO 1 is the forward primer sequence for the 4.7kb fragment from pRH004 (Table 6).

SEQ **ID** NO 2 is the reverse primer sequence for the 4.7kb fragment from pRH004 (Table 6).

SEQ **ID** NO 3 is the forward primer sequence for the 2.lkb fragment from pRH004 (Table 6).

SEQ **ID** NO 4 is the reverse primer sequence for the 2.lkb fragment from pRH004 (Table 6).

SEQ **ID** No 5 is the forward primer sequence for the Venus gene with flexible **GSio**-linker (Table 6).

SEQ **ID** No 6 is the forward primer sequence for AtSeipin2 (Table 6).

SEQ **ID** No 7 is the reverse primer sequence for AtSeipin2 with flexible **GSio**-linker (Table 6).

SEQ ID No 8 is the forward primer sequence for AtSRPl (Table 6).

SEQ ID No 9 is the reverse primer sequence for AtSRPl with flexible **GSio**-linker (Table 6).

SEQ ID No 10 is the forward primer sequence for ppOIel (Table 6).

SEQ ID No 11 is the reverse primer sequence for ppOIel with flexible **GSio**-linker (Table 6).

SEQ ID No 12 is the forward primer sequence for PP1S325 (Table 6).

SEQ ID No 13 is the reverse primer sequence for PP1S325 with flexible **GSio**-linker (Table 6).

SEQ ID No 14 is the forward primer sequence for FiblA (Table 6).

SEQ ID No 15 is the reverse primer sequence for FiblA with flexible **GSio**-linker (Table 6).

SEQ ID No 16 is the forward primer sequence for FiblB (Table 6).

SEQ ID No 17 is the reverse primer sequence for FiblB with flexible **GSio**-linker (Table 6).

SEQ ID No 18 is the forward primer sequence for PTS with flexible **GSio**-linker (Table 6).

SEQ ID No 19 is the reverse primer sequence for PTS (Table 6).

SEQ ID No 20 is the forward primer sequence for PTS only (Table 6).

SEQ ID No 21 is the reverse primer sequence for PTS only (Table 6).

SEQ ID No 22 is the forward primer sequence for Insert Check (Table 6).

SEQ ID No 23 is the reverse primer sequence for Insert Check (Table 6).

SEQ ID No 24 is the reverse primer sequence for ppOIel with rigid $EAAAK_3$-linker (Table 6).

SEQ ID No 25 is the forward primer sequence for PTS with rigid $EAAAK_3$-linker (Table 6).

SEQ ID No 26 is the sequence of the plasmid vector pRH004.

SEQ ID No 27 is the sequence of the plasmid vector with Venus fluorescence protein.

SEQ ID No 28 is the coding sequence of (-)-p-pinene synthase from Citrus limon (AF5 14288.1).

SEQ ID No 29 is the protein sequence for (-)-p-pinene synthase from *Citrus limon* (id=AAM53945.1)

SEQ ID No 30 is the coding sequence of (+)-limonene synthase 1 from *Citrus limon* (AF5 14287.1).

SEQ ID No 31 is the protein sequence for (+)-limonene synthase 1 from *Citrus limon* (id=AAM53944.1)

SEQ ID No 32 is the coding sequence of γ-terpinene synthase from *Citrus limon* (AF5 14286.1).

SEQ ID No 33 is the protein sequence for γ-terpinene synthase from *Citrus limon* (id=AAM53943.1)

SEQ ID No 34 is the coding sequence of S-linalool synthase (LINS) from *Coriandrum sativum* (KF700700.1).

SEQ ID No 35 is the protein sequence for S-linalool synthase (LINS) from *Coriandrum sativum* (id=AHC54051.1)

SEQ ID No 36 is the coding sequence of 1,8-cineole synthase from *Salvia officinalis* (AF05 1899.1).

SEQ ID No 37 is the protein sequence for 1,8-cineole synthase from *Salvia officinalis* (id= AAC26016.1).

SEQ ID No 38 is the coding sequence of (+)-bornyl diphosphate synthase from *Salvia officinalis* (AF05 1900.1).

SEQ ID No 39 is the protein sequence for (+)-bornyl diphosphate synthase from *Salvia officinalis* (*i&=AAC26*017*.1)*

SEQ ID No 40 is the coding sequence of sabinene synthase from *Salvia officinalis* (AF051901.1).

SEQ ID No 41 is the protein sequence for sabinene synthase from *Salvia officinalis* (id= AAC26018.1).

SEQ ID No 42 is the coding sequence of ocimene synthase (EBOS) from *Lotus corniculatus var. japonicus* (AY575970.1).

SEQ ID No 43 is the protein sequence for ocimene synthase (EBOS) from *Lotus corniculatus var. japonicus* (id=AAT86042.1)

SEQ ID No 44 is the coding sequence of fenchol synthase (FES) from *Ocimum basilicum* (AY693648.1).

SEQ ID No 45 is the protein sequence for fenchol synthase (FES) from *Ocimum basilicum* (id=AAV63790.1)

SEQ ID No 46 is the coding sequence of borneol synthase from *Chammaecyparis obtusa* (AB120957.1).

SEQ ID No 47 is the protein sequence for borneol synthase from *Chammaecyparis obtusa* (id=BAC92722.1)

SEQ ID No 48 is the coding sequence of cadinene synthase (TPS) from *Cucumis melo* (NM_001297453.1).

SEQ ID No 49 is the protein sequence for cadinene synthase (TPS) from *Cucumis melo* (id=NP 001284382.1)

SEQ ID No 50 is the coding sequence of Patchoulol synthase from *Pogostemon cablin* (KP694233.1).

SEQ ID No 51 is the protein sequence for Patchoulol synthase from *Pogostemon cablin* (id=ALQ43502.1)

SEQ ID No 52 is the coding sequence of valencene synthase (TPS1) from *Citrus sinensis* (NM_001288856.1).

SEQ ID No 53 is the protein sequence for valencene synthase (TPS1) from *Citrus sinensis* (id=NP_001275785.1)

SEQ ID No 54 is the coding sequence of a-bisabolol synthase from *Matricaria chamomilla var. recutita (*K*M259*07.1*)*.

SEQ ID No 55 is the protein sequence for a-bisabolol synthase from *Matricaria chamomilla var. recuti* (id=AIW00681.1)

SEQ ID No 56 is the coding sequence of alpha-humulene synthase from *Zingiber zerumbet zssl* (AB247331.1).

SEQ ID No 57 is the protein sequence for alpha-humulene synthase from *Zingiber zerumbet zssl* (id=BAG12020.1)

SEQ ID No 58 is the coding sequence of CYP71BA1 for P450 mono-oxygenase from *Zingiber zerumbet (A*B*33* 1234.1).

SEQ ID No 59 is the protein sequence for CYP71BA1 for P450 mono-oxygenase from *Zingiber zerumbet* (id=BAJ39893.1)

SEQ ID No 60 is the coding sequence of short-chain dehydrogenase/reductase 1 from *Zingiber zerumbet zsdl* (AB480831.1).

SEQ ID No 61 is the protein sequence for short-chain dehydrogenase/reductase 1 from *Zingiber zerumbet zsdl* (id=BAK09296. 1)

SEQ ID No 62 is the coding sequence of santalene synthase from *Santalum album* (HQ343276.1).

SEQ ID No 63 is the protein sequence for santalene synthase from *Santalum album* (id=ADO87000.1)

SEQ ID No 64 is the coding sequence of CYP736A167 from *Santalum album* (KU169302.1).

SEQ ID No 65 is the protein sequence for CYP736A167 from *Santalum album* (id=AMR44 190.1)

SEQ ID No 66 is the coding sequence of geraniol synthase from *Citrus jambhiri RlemTPS3* (AB691531.1).

SEQ ID No 67 is the protein sequence for geraniol synthase from *Citrus jambhiri RlemTPS3* (id=BAM29049.1)

SEQ ID No 68 is the coding sequence of geraniol dehydrogenase (GEDH) from *Ocimum basilicum* (AY879284.1).

SEQ ID No 69 is the protein sequence for geraniol dehydrogenase (GEDH) from *Ocimum basilicum* (id=AAX83 107.1)

SEQ ID No 70 is the coding sequence of acetyl CoA geraniol/citronellol acetyltransferase (AAT1) from *Rosa hybrid cultivar* (AY850287. 1).

SEQ ID No 71 is the protein sequence for acetyl CoA geraniol/citronellol acetyltransferase (AAT1) from *Rosa hybrid cultivar* (id=AAW3 1948. 1)

SEQ ID No 72 is the coding sequence of predicted protein (PHYPADRAFT _80169) (also known as PpOlel) from *Physcomitrella patens subsp. patens* (XM_001766345.1).

SEQ ID No 73 is the protein sequence for predicted protein (PHYPADRAFT _80169) (also known as PpOlel) from *Physcomitrella patens subsp. patens* (id=XP_00 1766397.1)

SEQ ID No 74 is the coding sequence of Pp3c8_820Vl.l (also known as PpSeipinlS325) from *Physcomitrella patens subsp. patens.*

SEQ ID No 75 is the protein sequence for Pp3c8_820Vl.l (also known as PpSeipinlS325) from *Physcomitrella patens subsp. patens.*

SEQ ID No 76 is the coding sequence of Putative adipose-regulator}' protein (Seipin) (also known as AtSeipin 2) from *Arabidopsis thaliana* (NM_102716.4).

SEQ ID No 77 is the protein sequence for Putative adipose-regulatory protein (Seipin) (also known as AtSeipin 2) from *Arabidopsis thaliana* (id=NP_l 74269. 1)

SEQ ID No 78 is the coding sequence of Rubber elongation factor protein (REF) (also known as At SRP1) from *Arabidopsis thaliana* (NM_179525.3).

SEQ ID No 79 is the protein sequence for Rubber elongation factor protein (REF) (also known as At SRPl) from *Arabidopsis thaliana* (id=NP_849856.1).

SEQ ID No 80 is the coding sequence of Fibrillin (FIB) (also known as AtFibrillin 1A) from *Arabidopsis thaliana* (NM_1 16640.5).

SEQ ID No 81 is the protein sequence for Fibrillin (FIB) (also known as AtFibrillin 1A) from *Arabidopsis thaliana* (id=NP_1923 11.1).

SEQ ID No 82 is the coding sequence of Plastid-lipid associated protein PAP/fibrillin family protein (also known as At Fibrillin IB) from *Arabidopsis thaliana* (NM_l 18350.3).

SEQ ID No 83 is the protein sequence for Plastid-lipid associated protein PAP/fibrillin family protein (also known as AtFibrillin IB) from *Arabidopsis thaliana* (id=NP_193955.1).

SEQ ID No 84 is the sequence for the GS-linker (Table 6).

SEQ ID No 85 is the sequence for the Rigid linker (Table 6).

SEQ ID No 86 is the sequence for the LP4/2A linker (Table 6)

REFERENCES

Bach SS *et al.* Heterologous stable expression of terpenoid biosynthetic genes using the moss *Physcomitrella patens.* In: *Methods in Molecular Biology* Edited by Rodriguez-Concepcion M, vol. The Plant Isoprenoids; 2013: Springer, New York, USA, ISBN: 978-1-4939-0605-5.

Bach SS *etal.* "Heterologous stable expression of terpenoid biosynthetic genes using the moss Physcomitrella patens." In: *Methods in Molecular Biology* Edited by Rodriguez-Concepcion M, vol. 1153, Plant Isoprenoids; Methods and Protocols, Methods in Molecular Biology, 2014.

Chen F *et al.* The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J* 201 1 , 66(l):212-229.

Chiyoda *et al.* Direct transformation of the liverwort Marchantia polymorpha L. by particle bombardment using immature thalli developing from spores. *Plant Cell Reports* 2008, 27: 1467;

Cove DJ *et al.* The moss *Physcomitrella patens:* a novel model system for plant development and genomic studies. In: *Emerging Model Organisms: A Laboratory Manual,* Cold Spring Harbor Protocols, Vol. 1. CSHL Press, Cold Spring Harbor, N.Y., USA. 2009, 69-104.

Decker EL *et al.* "Current achievements in the production of complex biopharmaceuticals with moss bioreactors." *Bioprocess BiosystEng.,* 2008, 31(l):3-9;

Gallagher, Sean R. *et al.* "Quantitation of DNA and RNA with absorption and fluorescence spectroscopy." *Current Protocols in Human Genetics,* 2007: A-3D.

Hamberger B *et al.* Plant P450s as versatile drivers for evolution of speciesspecific chemical diversity In: *Phil Trans R Soc B* (in press). 2013.

Lopez-Obando *et al. :* "Simple and Efficient Targeting of Multiple Genes Through CRISPR-Cas9 in *Physcomitrella patens*." *G3: Genes, Genomes, Genetics* 2016, 6(11): 3647-3653;

King, Brian Christopher, *et al.:* "In vivo assembly of DNA-fragments in the moss, Physcomitrella patens." *Scientific reports,* 2016, 6:1-7

Knight *et al:* "Moss gene technology." *Molecular Plant Biology* 2002, 2, 285-301 .

Koprivova A *et al.* N- glycosylation in the moss *Physcomitrella patens* is organized similarly to that in higher plants. *Plant Biol* 2003, 5(6): 582-591;

Kubota *et al.:* "Efficient Agrobacterium-Mediated Transformation of the Liverwort Marchantia polymorpha Using Regenerating Thalli." *Bioscience, Biotechnology and Biochemistry* 2013, 77: 167-172).

Mega T. Plant-type N-glycans containing fucose and xylose in *Bryophyta* (mosses) and *Tracheophyta* (ferns). *Biosci Biotechnol Biochem* 2007, 7 1 (12):2893-2904).

Murphy, D. J. Plant lipids : biology, utilisation and manipulation. (Blackwell, 2005).

Rensing SA *et al.* Protein encoding genes in an ancient plant: analysis of codon usage, retained genes and splice sites in a moss, *Physcomitrella patens.* Bmc *Genomics* 2005, 6:43.;

Reski R: "Development, genetics and molecular biology of mosses." *Botanica Acta* 1998, 111: 1-15;

Reski, R: "Development, Genetics and Molecular Biology of Mosses." *Plant Biology*, 1998, 111(1): 1-15.

Reski R: "Molecular genetics of *Physcomitrella." Planta* 1999, 208: 301-309;

Reutter, K *et al.* "Production of a heterologous protein in bioreactor cultures of fully differentiated moss plants." *Plant Tissue Culture and Biotechnology*, 1996, 2(3): 142-147;

Schaefer DG *et al.* Efficient gene targeting in the moss *Physcomitrella patens. Plant J* 1997, 11 (6): 1195-1206.

Simonsen HT *et al.* Perspectives on using *Physcomitrella patens* as an alternative production platform for thapsigargin and other terpenoid drug candidates. *Perspect Medici n Chem* 2009, 3: 1-6.

Sugano *et al:* "CRISPR/Cas9-Mediated Targeted Mutagenesis in the Liverwort Marchantia polymorpha L." *Plant Cell Physiol* 2014, 55(3): 475-481;

Vietor, R *et al.* Protein N-glycosylation is similar in the moss *Physcomitrella patens* and in higher plants. *Planta* 2003, 218(2):269-275;

Yonekura-Sakakibara K. *et al.* An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plc»ıt J* 201 1, 66(1): 182-193.

Zhan X *et al.* Metabolic engineering of the moss *Physcomitrella patens* to produce the sesquiterpenoids Patchoulol and α/β-santalene. *Frontiers in Plant Science* 2014, 5: 636.

CLAIMS

1. A transgenic bryophytic cell capable of producing geranyl pyrophosphate (GPP), farnesyl-pyrophosphate (FPP) and/or geranylgeranylpyrophosphate (GGPP) comprising at least one heterologous nucleic acid molecule encoding at least one chimeric protein comprising at least a first and a second polypeptide where said first and second polypeptides are operationally linked and said first polypeptide is a lipid body-associated protein and said second polypeptide is a biosynthetic enzyme.

2. The cell according to claim 1, characterized in that the bryophytic cell is a moss cell.

3. The cell according to any of the preceding claims, characterized in that said moss cell is *Physcomitrella patens.*

4. The cell according to any of the preceding claims, characterized in that said lipid body-associated protein is selected from the group of oleosin, fibrillin, seipin, perilipin, small rubber particle protein 1, small rubber particle protein 2 and small rubber particle protein 3.

5. The cell according to any of the preceding claims, characterized in that said biosynthetic enzyme is selected from the group of monoterpene synthases, sesquiterpene synthases, cytochrome P450, alcohol dehydrogenase, acetyl transferase, aldehyde reductase, aldehyde dehydrogenase.

6. The cell according to any of the preceding claims, characterized in that said at least one chimeric protein encodes a third polypeptide being operationally linked to said first or said second polypeptide, wherein said third polypeptide is a biosynthetic enzyme.

7. The cell according to any of the preceding claims, characterized in that said cell comprises at least a first and a second heterologous nucleic acid molecule, said first heterologous nucleic acid encoding a first chimeric protein and said second heterologous nucleic acid molecule encoding a second chimeric protein, where both said first chimeric protein and said second

chimeric protein encode at least a first polypeptide being a lipid body-associated protein and a second polypeptide being a biosynthetic enzyme.

8. The cell according to any of the preceding claims, characterized in that said second polypeptide is a biosynthetic enzyme selected from the group of β-pinene synthase (EC 4.2.3.120), limonene synthase (EC 4.2.3.114), γ-terpinene synthase (EC 4.2.3.114), S-linalool synthase (EC 4.2.3.25), 1,8-cineol synthase (EC 4.2.3.108), bornyl diphosphate synthase (EC 5.5.1.8), sabinene synthase (EC 4.2.3.1 10), ocimene synthase (EC 4.2.3.106), fenchol synthase (EC 4.2.3.10), borneol synthase, cadinene synthase (EC 4.2.3.92), patchoulol synthase (EC 4.2.3.70), valencene synthase (EC 4.2.3.73), bisabolol synthase, α-humulene synthase (EC 4.2.3.104), CYP71BA1 (EC 1.14.13.150), zerumbone synthase (EC 1.1.1.326), santalene synthase (EC 4.2.3.82), CYP736A167, geraniol synthase (EC 3.1.7.11), geraniol dehydrogenase (EC 1.1.1.183) and geraniol acetyltransferase.

9. The cell according to any of the preceding claims, characterized in that said transgenic bryophytic cell comprises three different second polypeptide encoded by three different chimeric proteins where said three different second polypeptides are ot-humulene synthase (EC 4.2.3.104), CYP71BA1 (EC 1.14.13.150) and zerumbone synthase (EC 1.1.1.326) or said three different second polypeptides are geraniol synthase (EC 3.1.7.11), geraniol dehydrogenase (EC 1.1.1.1 83) and geraniol acetyltransferase.

10. The cell according to any of the preceding claims, characterized in that said transgenic bryophytic cell comprises two different second polypeptide encoded by two different chimeric proteins where said two different second polypeptides are santalene synthase (EC 4.2.3.82) and CYP736A167.

11. The cell according to any of the preceding claims, characterized in that the biosynthetic enzyme relates to enzymes involved in the conversion of compounds to terpenes and terpenoids.

12. The cell according to any of the preceding claims, characterized in that terpenes and/or terpenoids are produced by the cell.

13. The cell according to claim 12, characterized in that retainment of the terpenes and/or terpenoids in lipid bodies of the cell is increased.

14. A bryophyte plant or bryophytic tissue comprising transgenic bryophytic cells as described in any of the claims 1-13.

15. A method of producing a transgenic bryophytic cell comprising introducing into a bryophytic cell capable of producing geranylpyrophosphate (GPP), farnesyl-pyrophosphate (FPP) and/or geranylgeranylpyrophosphate (GGPP) at least one heterologous nucleic acid molecule encoding at least one chimeric protein comprising at least a first and a second polypeptide where said first and second polypeptides are operationally linked and said first polypeptide is a lipid body-associated protein and said second polypeptide is a biosynthetic enzyme.

16. A method for preparing terpenes and terpenoids in a transgenic bryophytic cell comprising the steps of:

a) Introducing into a bryophytic cell capable of producing geranyl pyrophosphate (GPP), farnesyl-pyrophosphate (FPP) and/or geranylgeranyl pyrophosphate (GGPP) at least one heterologous nucleic acid molecule encoding at least one chimeric protein comprising at least a first and a second polypeptide where said first and second polypeptides are operationally linked and said first polypeptide is a lipid body-associated protein and said second polypeptide is a biosynthetic enzyme,

b) culturing the transgenic bryophytic cell to express or overexpress said at least one chimeric protein, hereby enabling the transgenic bryophytic cell to produce terpenes and terpenoids,

c) isolating the terpenes or terpenoids produced from the transgenic bryophytic cell.

17. A method for preparing terpenes and terpenoids according to claim 16, characterized in that said biosynthetic enzyme is amonoterpene synthase or a sesquiterpene synthase.

18. A method according to claim 16 or 17, characterized in that retainment of the terpenes and/or terpenoids in lipid bodies of the cell is increased.

19. Use of a bryophytic cell as described in any of the claims 1-13 or a bryophyte plant as described in claim 14 for production of terpenes and terpenoids.

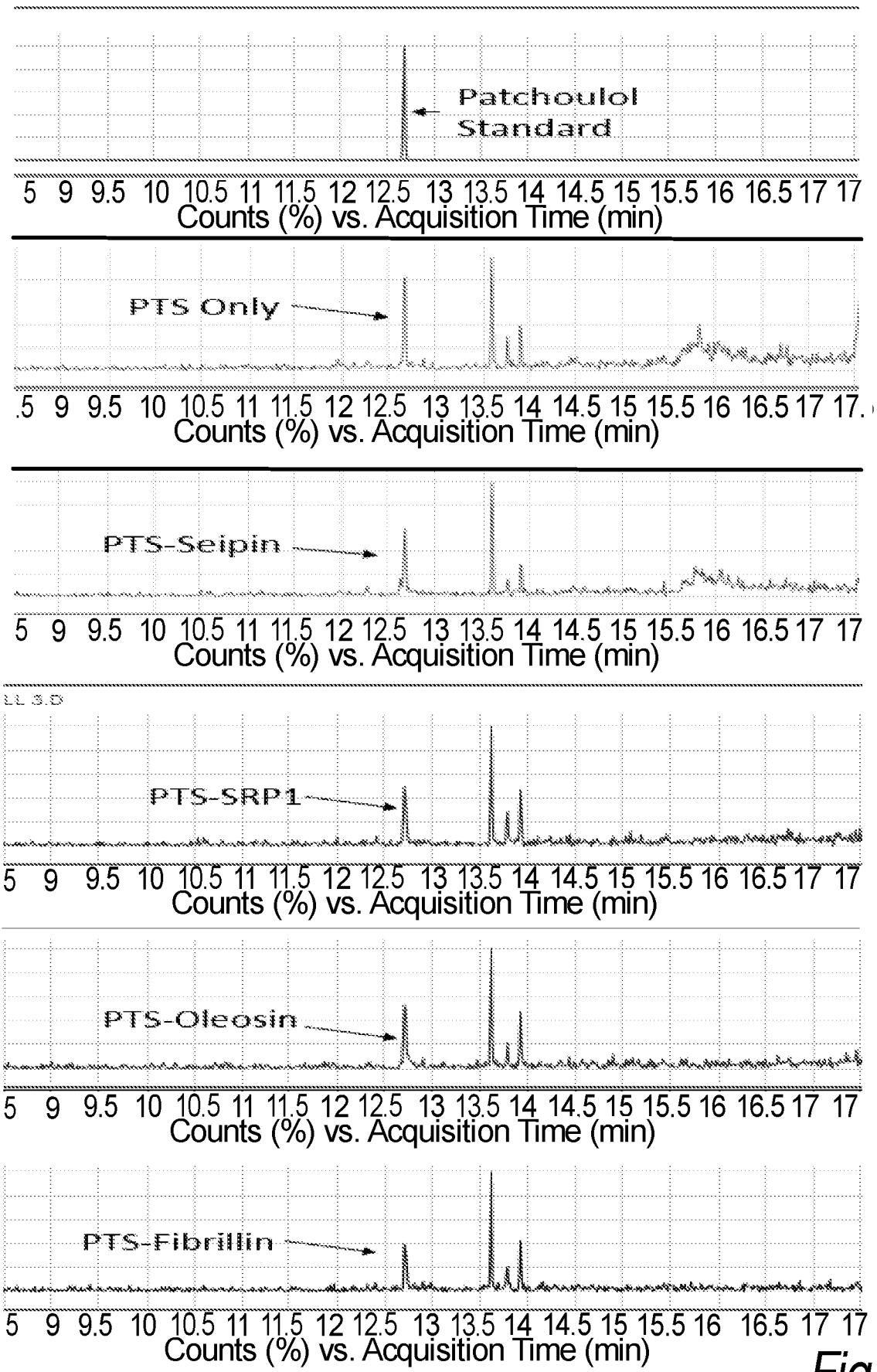
Patchoulol
Standard
5 9 9.5 10 10.5 11 11.5 12 12.5 13 13.5 14 14.5 15 15.5 16 16.5 17 17
Counts (%) vs. Acquisition Time (min)
PTS Only
.5 9 9.5 10 10.5 11 11.5 12 12.5 13 13.5 14 14.5 15 15.5 16 16.5 17 17.
Counts (%) vs. Acquisition Time (min)
PTS-Seipin
5 9 9.5 10 10.5 11 11.5 12 12.5 13 13.5 14 14.5 15 15.5 16 16.5 17 17
Counts (%) vs. Acquisition Time (min)
LL 3.D
PTS-SRP1
5 9 9.5 10 10.5 11 11.5 12 12.5 13 13.5 14 14.5 15 15.5 16 16.5 17 17
Counts (%) vs. Acquisition Time (min)
PTS-Oleosin
5 9 9.5 10 10.5 11 11.5 12 12.5 13 13.5 14 14.5 15 15.5 16 16.5 17 17
Counts (%) vs. Acquisition Time (min)
PTS-Fibrillin
5 9 9.5 10 10.5 11 11.5 12 12.5 13 13.5 14 14.5 15 15.5 16 16.5 17 17
Counts (%) vs. Acquisition Time (min)

Fig. 1

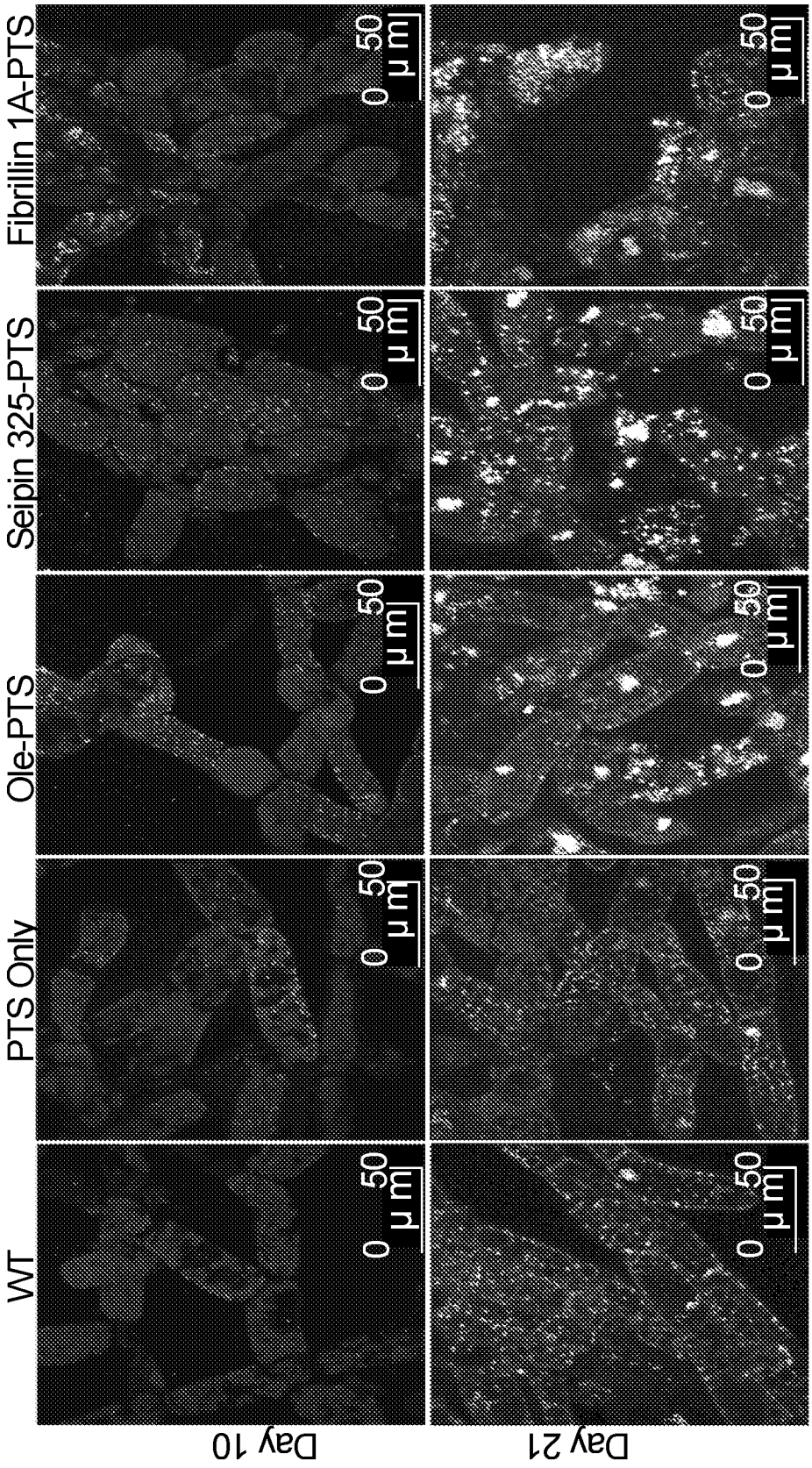
WT
PTS Only
Ole-PTS
Seipin 325-PTS
Fibrillin 1A-PTS
Day 10
Day 21
0 50 µm

Fig. 2

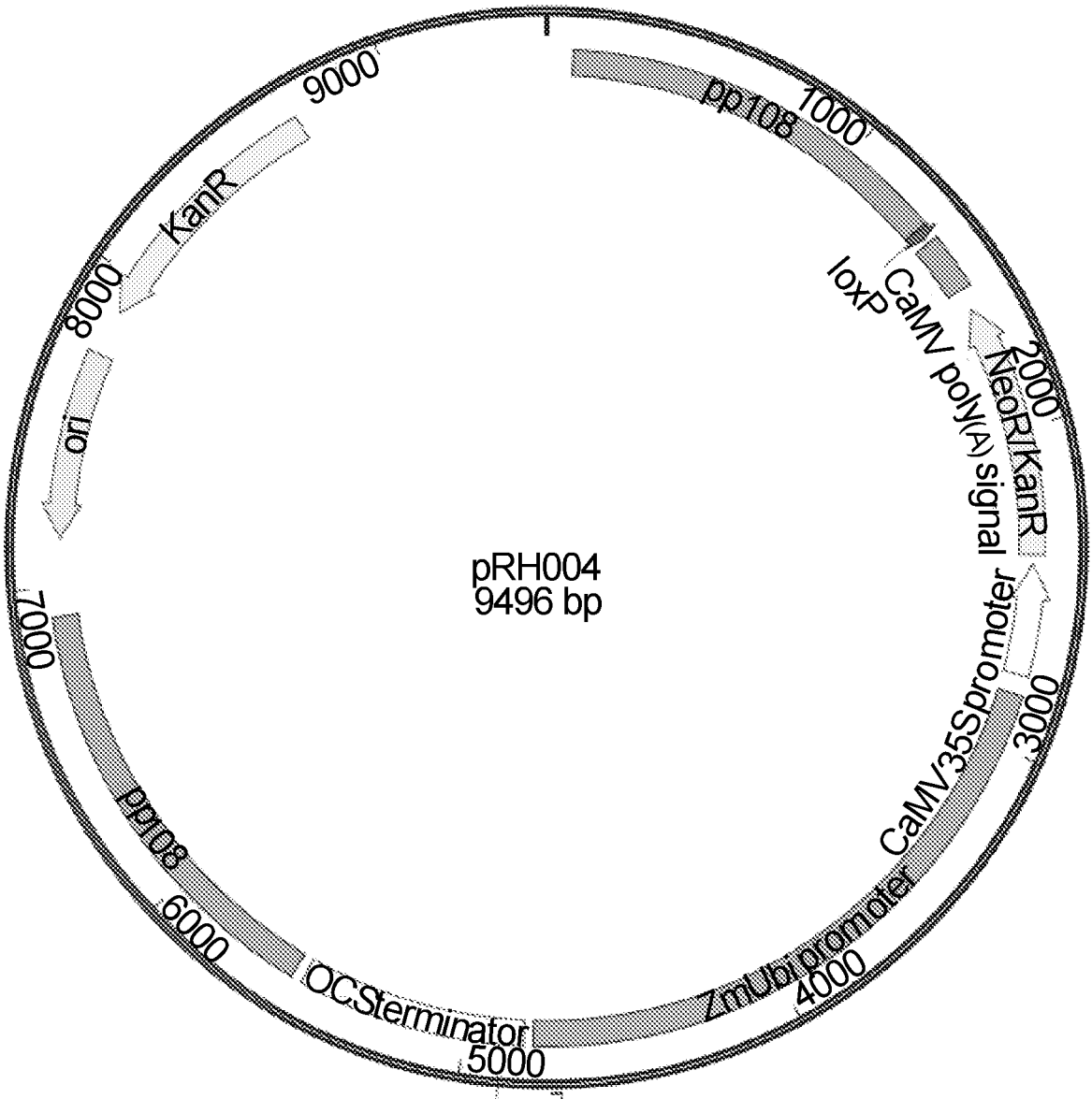
pRH004
9496 bp
1000
2000
3000
4000
5000
6000
7000
8000
9000
pp108
loxP
CaMV poly(A) signal
NeoR/KanR
CaMV35Spromoter
ZmUbi promoter
OCSterminator
pp108
KanR
ori
(4864 .. 4888) Insert Check R
Insert Check F (4707 .. 4733)

Fig. 3

4 / 8

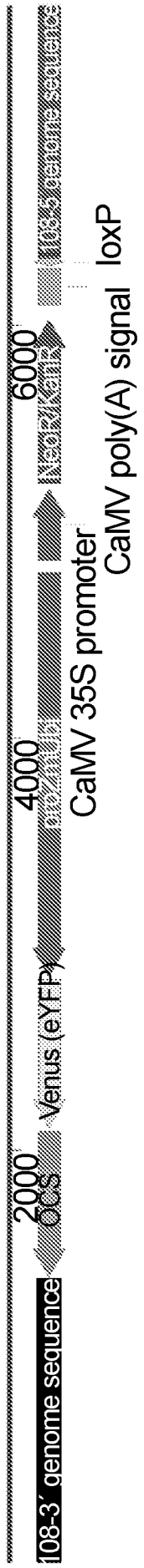
108-3´ genome sequence
OCS
Venus (eYFP)
nptII
CaMV 35S promoter
NeoRKanR
CaMV poly(A) signal
108-3´ genome sequence
loxP
2000
4000
6000

Fig. 4

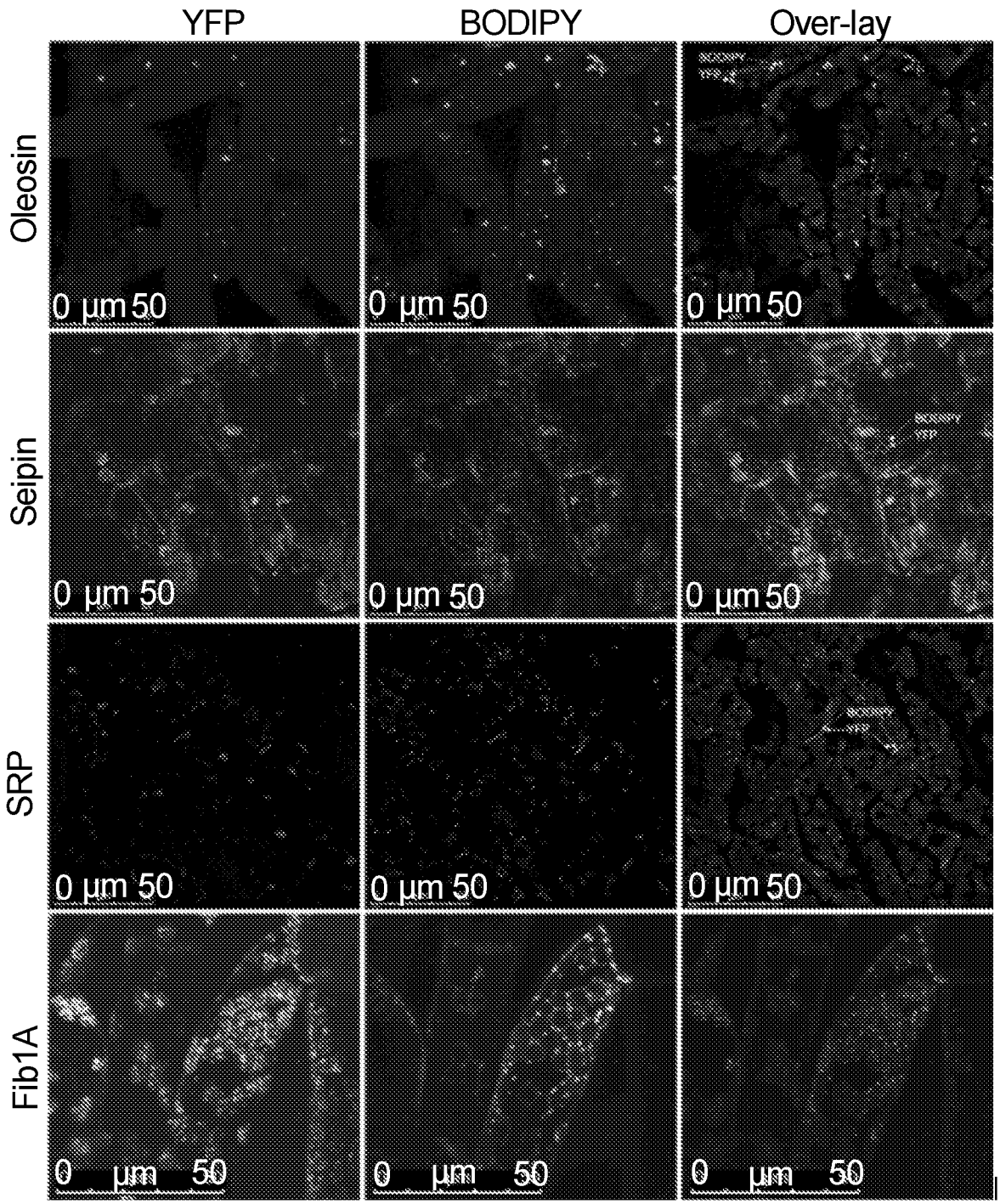

Fig. 5

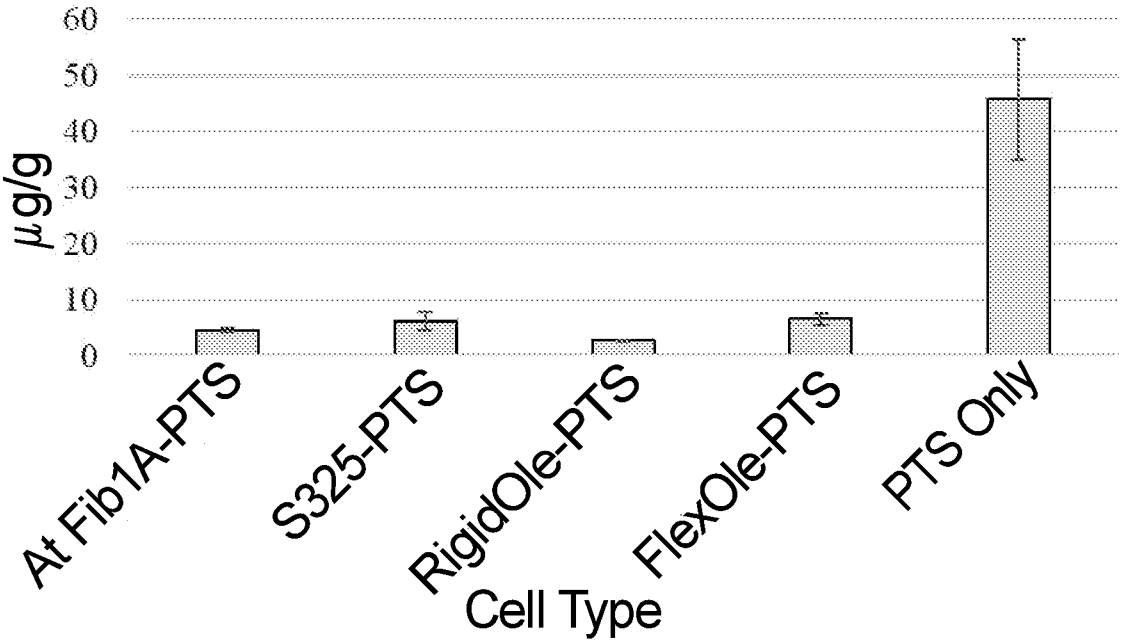
60
50
40
30
20
10
0
µg/g
At Fib1A-PTS
S325-PTS
RigidOle-PTS
FlexOle-PTS
PTS Only
Cell Type

Fig. 6a

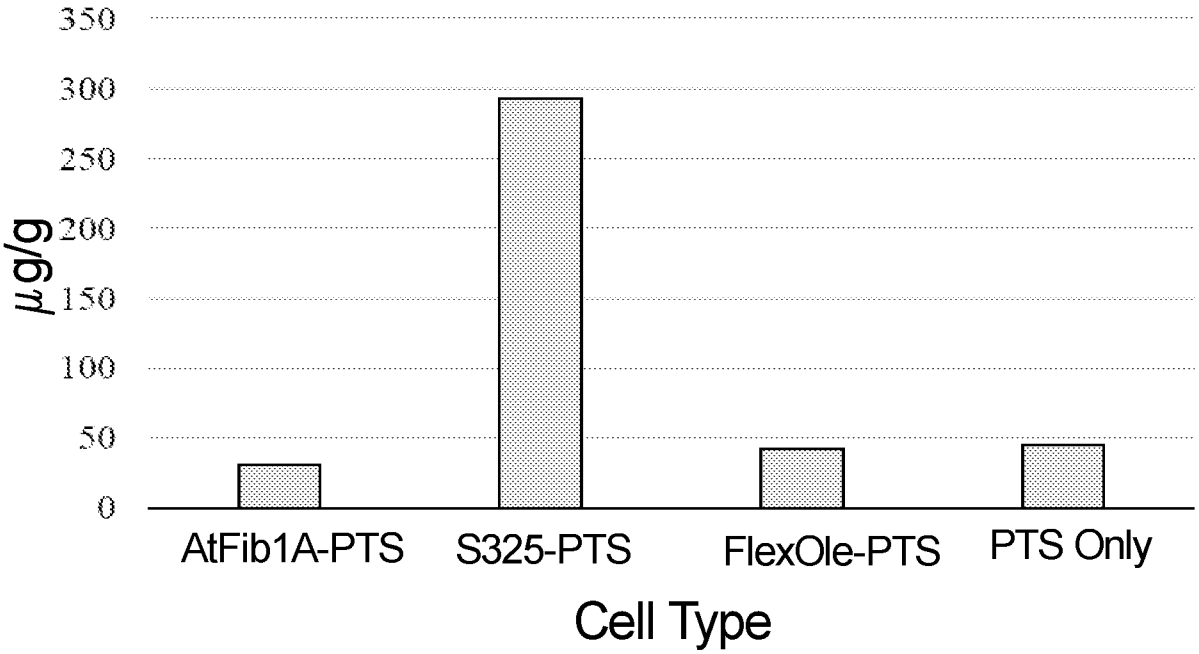
350
300
250
200
150
100
50
0
µg/g
AtFib1A-PTS
S325-PTS
FlexOle-PTS
PTS Only
Cell Type

Fig. 6b

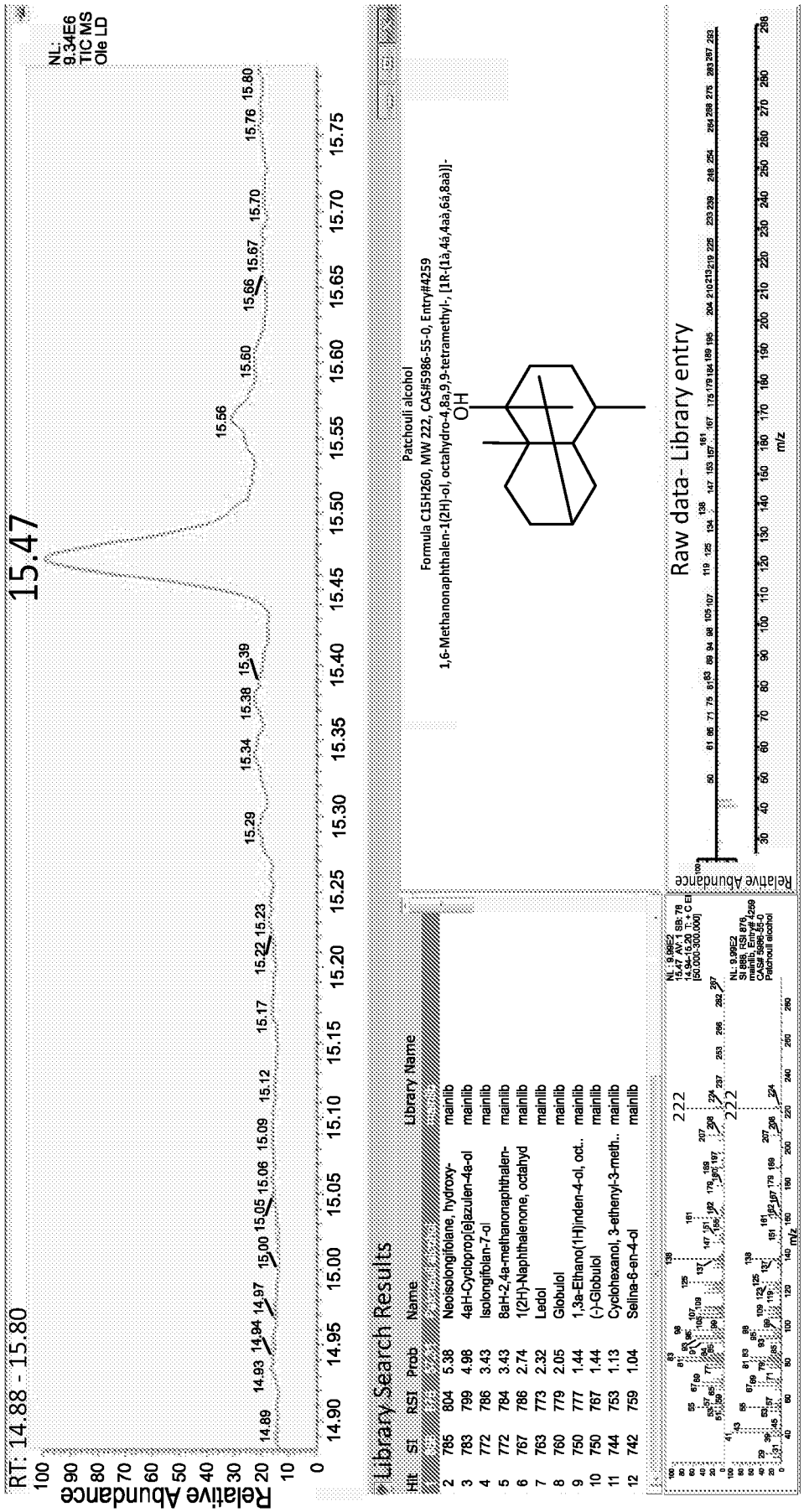
RT: 14.88 - 15.80
15.47
NL: 9.34E6 TIC MS Ole LD
Relative Abundance
Library Search Results
Hit ST RSI Prob Name Library Name
2 785 804 5.38 Neoisolongifolane, hydroxy- mainlib
3 783 799 4.98 4aH-Cyclopropa[e]azulen-4a-ol mainlib
4 772 786 3.43 Isolongifolan-7-ol mainlib
5 772 784 3.43 8aH-2,4a-methanonaphthalen- mainlib
6 767 786 2.74 1(2H)-Naphthalenone, octahyd mainlib
7 763 773 2.32 Ledol mainlib
8 760 779 2.05 Globulol mainlib
9 750 777 1.44 1,3a-Ethano(1H)inden-4-ol, oct.. mainlib
10 750 767 1.44 (-)-Globulol mainlib
11 744 753 1.13 Cyclohexanol, 3-ethenyl-3-meth.. mainlib
12 742 759 1.04 Selina-6-en-4-ol mainlib
Patchouli alcohol
Formula C15H26O, MW 222, CAS#5986-55-0, Entry#4259
1,6-Methanonaphthalen-1(2H)-ol, octahydro-4,8a,9,9-tetramethyl-, [1R-(1à,4á,4aà,6á,8aà)]-
OH
Raw data - Library entry
Relative Abundance
m/z

Fig. 7

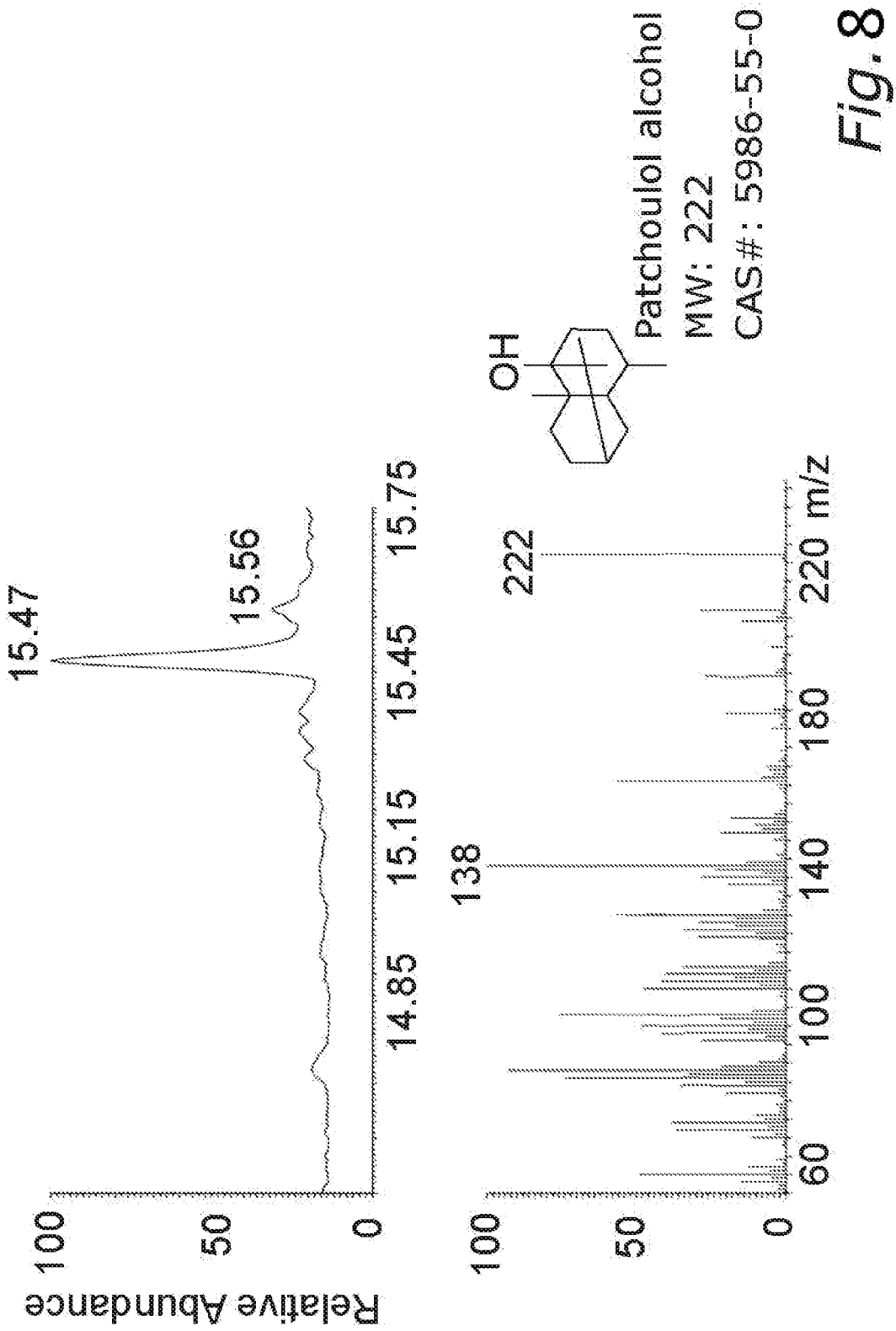
Relative Abundance
100
50
0
15.47
15.56
14.85
15.15
15.45
15.75
OH
Patchoulol alcohol
MW: 222
CAS#: 5986-55-0
222
138
60
100
140
180
220
m/z

Fig. 8

INTERNATIONAL SEARCH REPORT

International application No
PCT/DK2018/05Q231

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/88 C12N15/82 C12P7/Q4 C12P5/00
ADD. A01H11/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C07K C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal , WPI Data, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2011/048119 A2 (GEORG AUGUST UNI GOETTINGEN STI FTUNG OEFFENTLICHEN RECHTS [DE] ; FEUSSN) 28 Apri l 2011 (2011-04-28)	1-4, 14, 15
Y	page 7, paragraph 4 - page 9, paragraph 3 page 11, paragraph 2 page 12, paragraph 4 page 17, paragraph 2 -----	5-10, 16, 17, 19
Y	WO 2014/206412 Al (K0BENHAVNS UNI [DK] ; FI RMENICH & CI E [CH]) 31 December 2014 (2014-12-31) cited in the appl icati on page 3, line 24 - page 5, line 4 ----- -/--	5-10, 16, 17, 19

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) orwhich is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 October 2018	14/11/2018

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Bi l ang, Jürg

INTERNATIONAL SEARCH REPORT

International application No
PCT/DK2018/05Q231

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2003/126631 Al (MOLONEY MAURICE M [CA] ET AL) 3 July 2003 (2003-07-03) abstract paragraphs [0019] , [0051] , [0091] f i gure 1 -----	1-4,14, 15
A	Wo 2016/020689 Al (TEXAS A & M UNIV SYS [US] ; 0RNSTEIN JASON [GB]) 11 February 2016 (2016-02-11) the whole document -----	1-19

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/DK2018/05Q231

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2011048119 A2	28-04-2011	AU 2010309840 A1	19-04-2012
		CA 2777903 A1	28-04-2011
		DK 2491119 T3	12-09-2016
		EP 2491119 A2	29-08-2012
		PL 2491119 T3	30-12-2016
		US 2012240289 A1	20-09-2012
		WO 2011048119 A2	28-04-2011
WO 2014206412 A1	31-12-2014	US 2016222401 A1	04-08-2016
		WO 2014206412 A1	31-12-2014
US 2003126631 A1	03-07-2003	US 2003126631 A1	03-07-2003
		US 2004205844 A1	14-10-2004
WO 2016020689 A1	11-02-2016	EP 3177724 A1	14-06-2017
		US 2017226526 A1	10-08-2017
		WO 2016020689 A1	11-02-2016